

DOCUMENT RESUME

ED 161 146

95

EA 010 968

AUTHOR Cnudde, Charles F.
TITLE The Public Administration of the Public Schools:
Complex Policy Models of Educational Achievement.
Final Report.
INSTITUTION Michigan State Univ., East Lansing. Center for Public
Policy Analysis.
SPONS AGENCY National Inst. of Education (DHEW), Washington,
D.C.
BUREAU NO PRD-13111
PUB DATE 76
GRANT NEG-00-3-0206
NOTE 137p.; Chart on page 57 may be illegible

EDRS PRICE MF-\$0.83 HC-\$7.35 Plus Postage.
DESCRIPTORS *Academic Achievement; Educational Research;
Elementary Secondary Education; School
Responsibility; School Role; Socioeconomic
Background; *Socioeconomic Influences; Statistical
Analysis; Tables (Data)
IDENTIFIERS Coleman (James S); *Coleman Report

ABSTRACT

Previous researchers, notably James Coleman, found that schools had very weak effects on student achievement and that socioeconomic factors had very strong effects. They made this discovery by measuring the effects of schools when socioeconomic factors were held constant, and they found negligible effects of schools. This study's hypotheses are based on the suggestion that Coleman's analyses may have ignored the dependence of school factors on socioeconomic level of parents. The study tested policy models of educational achievement that include the following types of complexities: the dependence between school and background variables, nonlinear effects of one or more independent variables on performance, and differential levels of measurement error leading to differential bias in the estimates of effects from independent variables. To test the hypotheses, the Coleman data were used in linear regression equations in which achievement scores are dependent on different combinations of independent variables. The findings indicate that school characteristics are dependent on the socioeconomic backgrounds of the families of the students. This suggests that school effects may be much stronger than previously believed but may themselves depend on the socioeconomic backgrounds of their neighborhoods. This analysis suggests the need for a return to research on which aspects of school practices lead to higher achievement results. (Author/JM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED161146

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

The Public Administration of the Public Schools:
Complex Policy Models of Educational Achievement

"National Institute of Education Project No. ORD 13111"

Charles F. Cnudde

Center for Public Policy Analysis
Michigan State University
East Lansing, Michigan

1976

NEE-00-3-026

The research reported herein was performed pursuant to a grant contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

EA 010 968

TABLE OF CONTENTS

	<u>Page</u>
Summary	1
Introduction	1
Objectives	4
Review of Related Research	5
Procedures	27
Results	37
Conclusions	49
Recommendations	51
Bibliography	53
Appendices	55
Tables and Illustrations	
Figure 1	11
Table 1	34
Table 2	45
Table 3	47
Table 4	48
Appendix Tables	55

Summary

The primary findings of the study are that 1) characteristics of American schools closely follow differences in the "average" socio-economic background characteristics of the families of the children in those schools, 2) the estimated effect of family background upon student achievement is significantly reduced when school differences are "held constant", 3) controls for school differences reduce the effect on achievement of measures of learning practices in the family, 4) family learning practices and school differences together significantly reduce family background influences on achievement.

The primary conclusion is that previous research underestimated the importance of schools and overestimated the importance of family socio-economic background factors. The reason that these misinterpretations occurred is that the administrative organization of schools in the United States was largely ignored. The public policy in favor of a decentralized school system means that we should expect a lack of independence between school characteristics and family background. Recognition of this fact leads to theoretical models that support a more complex but appropriate analysis of the importance of school differences on achievement.

Non-linear and non-additive effects were not found to have significant support. Further results suggest the anticipated effects of random measurement error: The results gain in consistency with increasing school grades and with increasing precision of subject matter achievement tests. The failure of tests of statistical significance in the previous literature can be shown to derive from differential levels of random measurement error in independent variables.

The results discussed in the body of the report are brief summaries of complex analyses which are reported upon in more detail in the appendices.

Introduction

The past decade of social science research on education in the United States would lead educational decision makers astray if they tried to use it for policy guidance. (See especially Armor, 1972; Coleman, 1966; Coleman, 1976; Jencks, 1972; and Mosteller and Moynihan, 1972.) That research seems to conclude that school differences play no role in important measures of school performance like student achievement. As a result one might advise decision makers that information on what takes place in the school is unimportant and that what one should look to in this area is family background such as the parental occupational and income factors. One might also advise that research on what aspects of the school and teacher performance are most important for achievement doesn't matter either and that the investment of public resources in this field would be less than optimal.

While their conclusions are the current "wisdom" among researchers, it appears that they have been drawn prematurely. The reason for saying this is that the previous research has been concerned with testing a simplistic model of educational effects. In the research reported here more complex models are evaluated. This evaluation leads to the conclusion that school factors are of central importance in achievement differences because of the consistent support for a more complex view of how schools fit into the overall local community.

That is to say the previous research did not take into consideration the fact that the schools are our primary example of a decentralized governmental public service system in the United States. Because of this decentralization we expect a close "fit" between the nature of the neighborhood of a school and

the type of school program that takes place. From this perspective we see that social science measures of the family background of the children in the school is very close to a sampling of the nature of the neighborhood of the school. When the previous researchers attempt to hold "constant" family background in their studies they ignore the important political fact of life of how we organize our schools. They therefore come to inappropriate conclusions.

The present research follows a different plan. It sets out a statistical model that is consistent with these political realities. It therefore achieves a test of the actual role of the school-by-school differences captured in the data used in previous research. It therefore leads to a return to research on how schools can maximize achievement results.

In summary it is with the intention of guiding future policy making that a more complex model has been examined. The model has been developed in response to the outpouring of recent studies that has fairly well demolished our previous, rather naive policy models. This model is similar to what Armor (1972) has called the "integration policy model." In it student background factors and school practices were thought to make unique contributions to student performance on achievement tests. The demolition of this model may mean that the schools have no impact and that therefore public policies designed to change school practices may have little impact on important social concerns like 1) racial equality, 2) social mobility, and 3) upgrading the nation's pool of talent.

Before we can come to this conclusion several models which include a role for school inputs should be first evaluated. Each of these models specifies an effect from school differences which is more complex than that included in our earlier thinking. Moreover, each of these models is consistent with the findings presented by research on these matters to date. Yet none have been

thoroughly subjected to empirical analysis. In this sense the previous analyses have accomplished an important and necessary first step; the debunking of a naive model. The second step is that reported here: an examination of what models do hold in these data. The second step appears necessary if social science research in this area is to inform policy making.

Objectives

The objectives of the study are to test policy models that include the following types of complexities:

- 1) collinearity or lack of independence between school and background variables;
- 2) non-linear effects of one or more independent variables upon performance;
- 3) differential levels of measurement error leading to differential bias in the estimates of effects from independent variables.

Each of these complexities have been tested with statistical models based upon prior theoretical specification of the model. The complexity that seems most consistent with the data based upon these tests is due to major departures from independence among supposedly independent variables compounded with severe measurement error. The results of the tests are referred to in the report. The bulk of the report includes the theoretical specification and tests of the complexities that show primary importance for American education as measured.

REVIEW OF RELATED RESEARCH

"The pathbreaking quality of EEOR had to do with its analysis of the relation of variation in school facilities to variation in levels of academic achievement. It reported so little relation as to make it almost possible to say there was none."

(Mosteller and Moynihan, 1972: 8)

Pettigrew and Green (1976) show several methodological difficulties in the previous research on education. Despite limitations of this type, the most important difficulty in this research seems to be theoretical or conceptual. If we wish to deal with important policy problems we must conceptualize in terms of models. Once we do that we see that the previous research assumed rather simple models. We are then lead immediately to the possibility that the introduction of a little more complexity might lead us to different conclusions.

The answers to research questions have been generated by the nature of the question:

What complex models account for the manner in which the social backgrounds of pupils interrelate with policy "manipulable" school differences to produce differences in measured achievement? To date questions about the effects of school inputs have been asked with rather simple statistical models (at least implicitly) in mind. As the above quote shows, the answers provided with these models have been almost uniformly discouraging from a public policy perspective. School-by-school differences seem to have little impact upon performance that is independent of family and other social background characteristics. This conclusion holds up whether the studies involve cross-sectional (Coleman, et al., 1966; Coleman, 1970; Jencks, 1972) or before and after designs (Armor, 1972). The great contribution of the burden of the evidence has been to make us rethink some of the simplistic assumptions that

"liberals" and others too quickly adopted, the notion that the schools could easily solve all of our social ills. Against this background of untried assumptions recent research has provided a healthy corrective which sets the stage for new investigations.

The research problem in the investigation therefore is whether models which go beyond our previous simplistic notions can help explain achievement in the public schools. For example, David Armor tested most of the assumed relations in the "integration policy model" which dominated policy thinking with respect to schools. He felt that much of the model needs serious revision. Our program of research begins by altering the model in ways that are unlike the assumptions of the ordinary statistical procedures that have borne the analyses to date.

In his thoughtful summary of the model as it applies to integration, Armor says, "Segregation and inequality combine to cause psychological damage in children resulting in lower achievement, lower aspirations, and less self-esteem." (Armor, 1972, p. 96). Now background characteristics associated with inequality can combine with school segregation in several ways. Our usual thinking, however, has hypothetically limited the ways that these factors combine. That is, we ask, given the scholastic disadvantages associated with poverty, the lack of books in the home, and a history which does not emphasize education; what additional effects are produced by a poor quality education (to the extent that we can measure this.) Thus, the very sophisticated procedures in chapter three of the Coleman Report (Coleman, et al., 1966) tell us how much school differences add to the variance explained in achievement over and above that explained by background characteristics. The Jencks study makes the hypothesized causal relations explicit; it also is essentially based upon the regression model of ordinary least squares statistical procedures. The point is not that these procedures were inappropriate to the theoretical model in our policy assumptions. In fact they were perfectly reflective of these assumptions. As a result of these tests

we can now reject that model. What is needed then is an alternative model.

For example, we think that political variables have one kind of impact on policy outcomes when the political battle involves distributive proposals and quite another impact when the proposals are less distributive. (Cnudde and McCrone, 1969). Similarly, the usual statistical model which assumes independent effects fails to come to grips with relations which interact in this way. The evidence that school differences do not add to our ability to explain performance in school over and above background characteristics does not provide sufficient ground to make conclusions about whether these factors combine as part of the same process.

The significance of an examination of more complex policy models seems clear. Policy makers in American educational systems can manipulate variables connected with school differences but they cannot change the background characteristics of students. This point is basic to one of the criticisms of policy-oriented research in American education. (Cain and Watts, 1970). Rejection of the assumptions in the original policy model in this field might cause policy makers to feel that there is little that social science research can contribute to the solution of their current problems. Before reaching this conclusion, however, it is necessary to investigate slightly more complex models which are different from those originally posited but which have theoretical support nevertheless. If other more complex models are consistent with the empirical evidence, then American education can be advised in ways to benefit millions of students (as well as the rest of us.) Current social science evidence indicates that we can reject our previous simplistic thinking about education. The conclusion that social science is unable to guide educational policy making seems premature.

What is needed in order to gain a new approach to this problem is to compare the findings of the previous literature in education research with the research in the field of public administration. If we take the Coleman report (1966), the study referred to as the EEOR in the quote from Mosteller and Moynihan, as the current wisdom then those conclusions are inconsistent with what one would expect from a public administration perspective. In the following pages we turn to the development of a theory of what one might expect of the evidence on student achievement given the previous research in these two fields. In the course of this theoretical specification we also consider a brief case study which further explicates and illustrates features of the theory. Only then are hypotheses drawn from the theory evaluated with the data from EEOR.

In the United States the trend in the administration of public services since the 1930's has been toward larger organizational units. (For a review and critique of this trend, see Ostrom, 1973.) We have had new Federal programs emerge. We have sometimes had the nationalization of many State and Local functions by Federal Agencies. We have attempts at consolidation of local governmental services; sometimes successful, sometimes not. At the State-level plans frequently have arisen to reduce the number of executive departments into a few "super" agencies under the control of the Governor. The logic underlying all of these proposals is quite similar: an increase in efficiency and control in order to achieve some policy goal.

Public education is a policy area that in large part appears as a "hold out" against this trend. There are Federal programs in education, to be sure. There are also State-level educational activities on the increase in many jurisdictions. In addition school district consolidation has led to larger organizational unity in many rural areas. However, the administrative authority in public education, especially in elementary and secondary programs, remains an unusually local matter:

School budgets are largely determined in response to local district vote. Staffing and coordination usually are in the hands of a local superintendent who establishes the structure for reporting on the delivery educational services. The superintendent ordinarily is legally responsible to an elected local board which has overall authority for planning and directing education programs. (Van Miller, et al., 1972.) In short the public schools tend to be organized on the basis of "place" whereas much of current public administration involves other than "place" as a bases of organization, to use Luther Gulick's categorization. (Gulick, 1973.)

Therefore, public education may serve as an interesting case study of how organization by place functions and therefore of the value of Gulick's categories for understanding current administrative organization. In addition, if we assume the applicability of his categories we may be in a position to further explicate their underlying logic through contrasting this policy area with others following different organizational fabrics. From such an explication we may derive testable propositions about the delivery of services in this policy area.

In his famous article on the "Proverbs" of administration, Herbert Simon pointed out the internal inconsistencies in the various bases of organization. (Simon, 1946.) If we set up purpose organizations we get one set of results and if we establish process agencies we get another. Each set of outcomes has advantages and disadvantages such that the good points of one are the other side of the coin: the bad points of the other. Because you have to make a choice over which type of organization that can be instituted (if you have organization by place you can't have organization by process) they are mutually exclusive. Because they are mutually exclusive and the good points of one are the bad points of the other, they seem logically inconsistent.

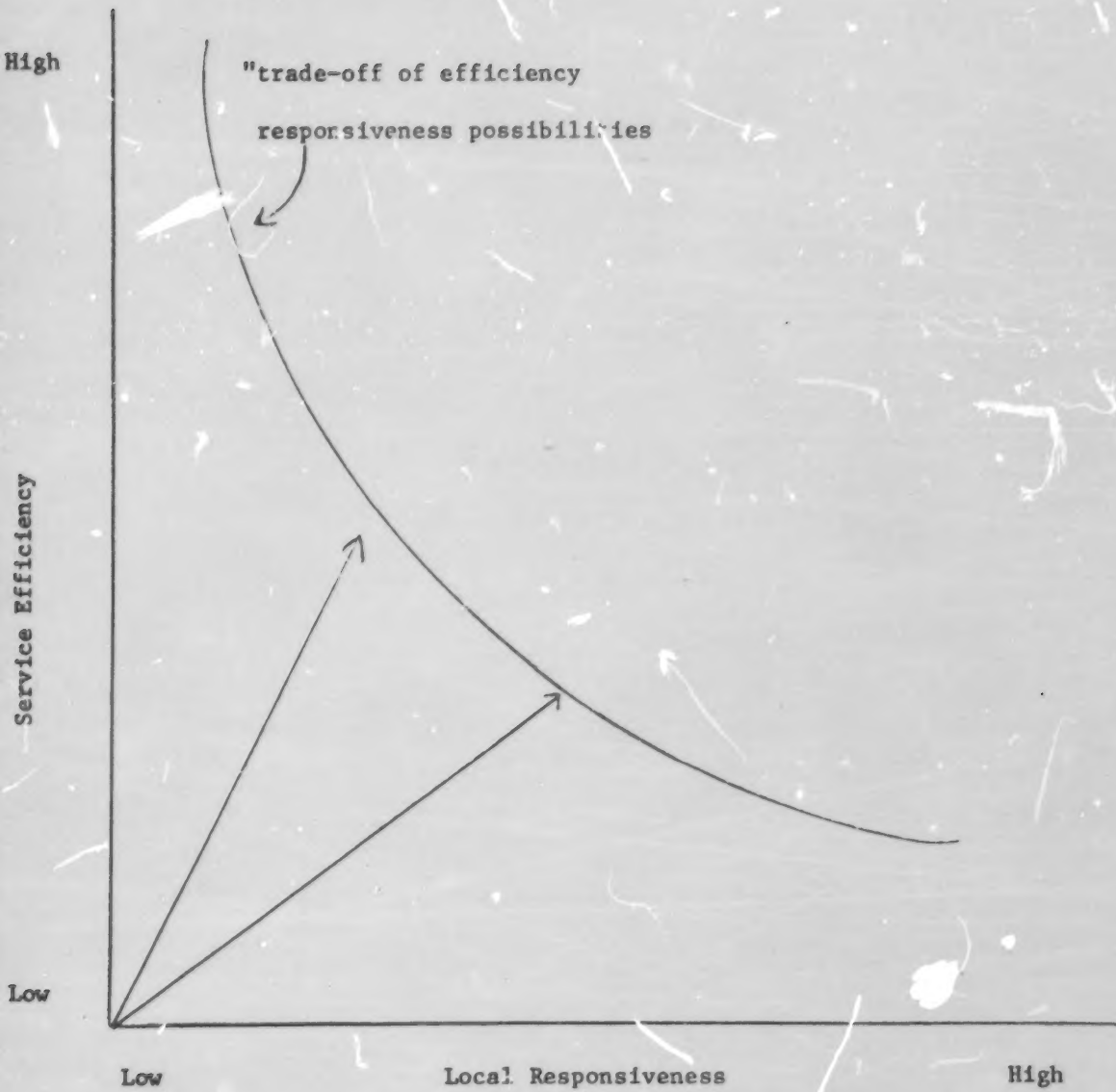
Because they are mutually exclusive and logically inconsistent the thought underlying the bases of organizing administrative work can not provide clear guidance on what choices to make in order to achieve a given organizational

goal such as efficiency in the delivery of public services. As a result the principles of organization set out by Gulick are at best "pre-theory" and perhaps even no more than proverbs of administration.

It seems to me that one might increase the predictive power of Gulick's principles of organization and thereby increase their theoretical status by looking further into the inconsistencies that Simon addressed. The usual conceptual way to treat apparent logical inconsistencies is to think of mutually exclusive options. Thus if you chose A you can't have B, and so forth. If the options are not discrete but vary by degree then we have a somewhat more sophisticated version of the same situation: the more A you chose, the less B you can have. In this manner we can conceive of the ratio of the degrees of A to B that are possible alternatives as a trade-off curve that displays the joint possibilities. Since more of one implies less than the other this curve would have a negative slope.

For example, we can think of the goals of responsiveness to the local community and efficiency in delivering a stated service in this manner. The first is increased through organization by place and the second is increased through purpose organization according to Gulick. On the other hand place organization will tend to decrease the efficiency and purpose organization will decrease responsiveness according to the usual treatment of Gulick's principles. Thus our example treats the inconsistency Simon referred to: both goals are valued and so we can't choose which type of organization to establish because their sets of outcomes counter balance each other.

On the other hand if we think of organization and goals at a higher conceptual level we can treat both as varying by degrees. We then have a trade-off curve as shown below. The curve shows that we can not simultaneously achieve high values of service efficiency and local responsiveness; high values of the one go hand-in-hand with low values of the other. Similarly two vectors, A and B, have been drawn to



intersect the trade-off curve at different points. A intersects at a point that corresponds to relatively high values on service efficiency and relatively low local responsiveness values. B intersects at a point with the opposite ranking of the two variables. Therefore A represents what we would conceptualize as consistent with a purpose organization effect and B with that from organization by place.

This conceptualization shows that there is in fact enough theoretical structure underlying the principles of organization notions to provide predictions about administrative outcomes. Consequently one can use the bases of organization logic to provide practical guidance on administrative options. For example, if you wish to maximize local responsiveness, and are willing to do so at a cost in efficiency, organize the agency on the basis of place. On the other hand if efficiency in providing the stated service is to be maximized even if doing so steps on the toes of citizens in certain areas, organize on the basis of purpose. It thus appears that Gulick's principles are inconsistent only when they are narrowly construed. When raised to a more abstract conceptualization we see that what we have is simply the usual situation where every benefit has its cost. The political process is then necessary to weigh the costs and benefits in order to make choices in a limited world.

From this perspective it is clear that our public education system was established in this country in order to maximize local responsiveness at a cost in the efficiency of the delivery of educational services. The outcome is as if we had decided that maximizing the educational achievement of our young people is not to be pursued regardless of other values. In fact one of the values held in higher regard is ensuring that the school as an administrative unit responds to the desires of the local community. If the community in some areas ranks football higher than algebra, than we have organized the schools to reflect

such preference ordering. On the other hand where the community ranks scholastic achievement above all else, the organizational bases of our schools helps to make sure that ranking is reflected in that community's schools. In short local control of schools is an organizational device that ensures a political relation between the nature of the local community and the services provided by the schools. Given differences across communities we should not be surprised then that schools vary in the degree to which their services include "good" education. Similarly we should not be surprised that not much education takes place at all in some schools.

The delegation of primary responsibility for the schools to local government is not the only aspect of "organization by place" that is important in the educational process. Within the local governmental units there is ordinarily a further decentralization of the educational organizations into "attendance units." The attendance unit is a geographical subdivision of a local school system which defines which areas of a community will send their children to which school. The attendance unit therefore normally consists of a school with its staff and top administrator and the region of the city, county, etc. that "belongs" to that school in the sense that the families in that region must send their children to that school. The "attendance unit" is the basis of the American concept of the neighborhood school and as such it stands for a very high degree of organization by place.

As a result we would not only expect school services to vary across local governments but also across neighborhoods within the same locality. We would especially expect to find this pattern in the United States because of the well-known homogeneity of our neighborhoods. Thus ordinarily we find residential choices being made via economic and other criteria which have the effect of sorting families and individuals into geographic units made up of other families and individuals quite like ourselves on those criteria. We then superimpose upon this situation a school system upon which includes a great deal of decentralization. We would

then expect a tendency for the individual schools to adapt to the neighborhood socio-economic make-up in the provision of services at least when there are socio-economic differences in the preferences on educational priorities.

To a certain extent there is evidence that entire neighborhoods take a proprietary interest in the going-ons of "their" school. However, we also know that in any type of political participation less than all qualified citizens become active enough to make their wishes known. Political scientists have pointed out for some time that political participation is a function of the incentives and means to act that citizens have. Citizens without the means to influence policy and without incentives to do so can not be expected to participate in actual decisions. The decentralization of the school system provides the means to influence decisions. But because everyone in the school's neighborhood doesn't have equally high incentives to act an effective school administrator doesn't always have to second-guess the wishes of the entire neighborhood when making decisions about educational services. Instead it would appear that a reasonable rule of thumb for an administrator would be to pay closest attention to that sub-set of the neighborhood or attendance unit that is most likely to become active-those who are most closely affected by changes in educational policy. This subset would be the direct consumers of the schools's services, the families of the students in the school itself.

Therefore, the socio-economic circumstances that are of most concern in school policy would be those of the families of the recipients of the school's services. If these families are not satisfied they have the greatest incentive to participate to change school services. The organization of the schools according to neighborhood gives them the political means-both formal and informal-to do so. With incentives and means to act it would be surprising if the schools did not adapt to the wishes to this sub-set of the citizenry.

A Case Study

An intensive case study of a neighborhood school over a two year period indicates that the participation bias can give rise to rather stable clientele groups even though schools as such are not organized on a clientele group basis. This observation raises a question of whether Gulick's categories are in fact mutually exclusive.

The case study was broken down into observation of policy and personnel issues. The former in fact was of the form of a "non-decision" in that there was a "mobilization of bias" in favor of an on-going educational policy. This "bias" excluded alternatives from active consideration and it seemed to control the outcomes of decisions that were of the form of actual consideration, including personnel decisions. The policy issue involved the set of educational objectives generally subsumed under the idea of the "open classroom."

The personnel issue was narrowed down to the decision to retain a particular probationary teacher. However, this decision was imbedded in a context of other personnel matters including the resignation of another teacher, general teacher morale in the school, and the replacement of the school principal.

For purposes of analysis the case may be treated chronologically. In September of a recent school year there was an assignment of a new third grade teacher to a small school in a relatively self-contained neighborhood of a nearly all white college town. The neighborhood was largely made up of academics. It therefore, had biases that favored scientific expertise in public decisions, social tolerance and political liberalism.

The teacher in question was a member of a minority group. As a result there was considerable community support for the teacher's presence in the school; support undoubtedly generated by the tolerance and liberalism of the neighborhood. One indicator of the high level of support was that the teacher was able to get away with behaviors that would not have been tolerated by non-minority group

members, including swearing at third grade students as a form of disciplinary action during the first week of school.

Direct observation of the classroom behavior of the teacher and the third grade children indicated that she utilized the "open classroom" style of instruction. Several of the hallmarks of this approach were in evidence. Children worked collectively in separate sub-groups on group projects much of the time. They were allowed to leave their desks to follow their own curiosity. They were not disciplined for interruptions of classmates or of the teacher. They were away from any possible assigned post for almost half of the time.

The "open classroom" approach was a matter of educational policy explicitly adopted by the principal and staff of the school in question. This approach furthermore was consistent with the policy biases of the parents in the neighborhood. The "open classroom" was a much discussed answer to several educational problems a few years ago. The approach would lead children to learn at their own pace rather than in an authoritarian "lock step." Consequently learning would be more effective because it would occur when the individual was ready for it and at the same time schools would avoid much of the repression of traditional approaches.

The enthusiasm for the "open classroom" stemmed from the tendency in the neighborhood toward a "modern" and "liberal" belief system. As one of the latest innovations to gain the support of educational "experts" the approach tied in with the attitudes of many parents toward the intelligent use of science and information in the solution of social problems. Furthermore, the stress on an open, democratic, non-repressive classroom atmosphere was consistent with the liberal orientation of much of the neighborhood. As a result the observed activity in the classroom was in style consistent with the policy of the school and that policy had the ideological support of many of the parents in the school's attendance area.

Another important value of the parents in the neighborhood stresses academic achievement. Many of the families in the attendance unit had at least one parent on a college or university faculty. Most of the balance of the area's population were attracted to the neighborhood largely because of the presence of the "good" school atmosphere identified with college-town neighborhoods. Thus in expressions observed at school meetings academic quality was a value underlying statements of approximately ninety percent of the parental activists, whether they were themselves academics or not.

A dispute over the retention of the teacher in question served as a focus for a series of events which brought the values of "technological liberalism" and academic excellence into conflict. First by mid-year a minority of parents who were previously supporters of the teacher began to complain to the teacher that there was a noted halt in the academic progress of their children, all of whom had previous records as high achievers. Eventually these complaints were taken to the principal. One of these parents in particular reports questioning why science had not been taught at all over the four month period. She further reports that the teacher first treated the question flippantly and when pressed said that the science "kits" had been ordered for some time but hadn't arrived in the classroom yet. Therefore, the class was emphasizing reading skills until the science materials arrived. This parent then went to the principal to ask why the science "kits" were so long in transit. She reports that the principal said that the "kits" had been available in the school since September. Why then didn't the teacher use them? The principal's answer to this question led the parent to feel that she was sticking her nose into a matter of school administration that should be left to the professionals.

Other parents among this minority had other complaints. One reports that she asked the teacher to stop using her son, the largest child in the class, as an enforcer of discipline and messenger because these jobs were taking too much of his classroom time. Despite her request she reports that these practices continued.

Another parent reports that he visited the classroom repeated to observe the teaching process. His specific complaint was that nearly forty percent of the school day was turned over to "free-time" on the days observed. Questioning of children in the class indicated that large portions of nearly every afternoon were turned over to "free-time" or "free-reading" it was alleged. Another parent reports a concern that English composition was not being introduced even though school board policy stated this subject as an objective for the grade. She claims that after repeated discussions in which she and the teacher disagreed over whether writing was being taught in the class, she finally gave up in frustration when she discovered that the teacher thought she was talking about penmanship. Still another parent reports that the allegation that the teacher failed to modify classroom practice when he complained that "free reading" did not constitute instruction in reading skills.

Most of these parents report experiences similar to the woman who complained about the lack of science instruction. When they went to the principal with their concerns, they say there was a lack of concrete response. They say they received general statements that the principal and staff were interested in academic achievement. Over and over again they report the principal told them that the goal of the school was to be the "best" school in the district, and already high achieving school system. Furthermore, the way to achieve this goal was to let the educational experts and authorities, like the principal and staff, work unhampered by "lay" people.

In the unfolding of these events several of the "complaining" parents organized to follow their concerns through the administrative process. After failing to obtain changes through direct discussions with the teacher and the teacher's superior, the principal, they met informally to see what else might be done. First they decided to find out how many other parents agreed with their complaints.

An attempt to encourage parents to attend the school's P.T.A. meetings to voice their concerns gave them such an indication. Some parents said they were also concerned by their children's lack of progress but that they did not want to move away from the "open classroom" in order to change the situation. Others focused more directly upon the "open classroom" and were satisfied with the teacher's performance. Other parents were enthusiastic of the teacher and put down criticisms as racially motivated. Still other parents didn't know there was a new teacher in the school and refused to go to P.T.A. meetings. Furthermore, discussions at meetings themselves were constrained by the presence of the staff at these meetings, including the principal and the teacher in question. Consequently it became clear that only a small minority of the parents in the attendance area were enough concerned about educational policy issues to state their preferences at a meeting. Of that small minority only from eight to ten parents were willing to raise educational achievement to such a high priority to want to take additional action with respect to the teacher in question.

Nevertheless, these parents say they decided to follow up their concern. During February they met with school district personnel officers, with the assistant superintendent for Instruction and finally in March a delegation met with the district superintendent. Prior to the meeting a fifteen page memorandum had been developed documenting their complaints about the educational quality of the instruction in the classroom in question. This material was submitted to the district authorities, including the Board of Education before the meeting with the superintendent.

At the meeting the parents summarized their criticisms as contained in their memorandum. At the conclusion of their statement the superintendent made his statement which included a commitment to "openness" and receptivity to parental preferences. On the other hand the superintendent balanced this position with a commitment to professional criteria in managing the school district. At the

conclusion of his statement the superintendent asked the parents what they wanted him to do about the situation.

Each of the parents present made an individual statement of what they personally wanted at this point. The statement of Mrs. Smith is representative of those made. Mrs. Smith, holder of an advanced degree in English

told of her own internal conflict in the situation.

She said "I don't feel comfortable challenging the decisions of the school district about what teachers to hire. In addition as a Christian I am a supporter of racial integration and affirmative action programs. So I am uncomfortable criticizing this particular teacher. But my most important desire of the School Board is the education of my children. My child in this classroom has received absolutely no instruction this year; in fact she is behind where she was last year at this time in some subjects. The rumor is that this teacher will be teaching the next grade next year so we are facing the possibility of two years that are simply lost in my child's education. What I want is to make sure my child doesn't have this teacher next year and from the point of education in general I don't think this teacher should be allowed in any classroom."

The superintendent responded by outlining the procedures used in making a decision to retain a first-year teacher. Then he said two things that hit the parents' committee like a bomb. He said that the decision had been made back in January to continue the teacher for a second probationary year and therefore the group was too late in communicating their concern. He also said that the only information about parental concern available when the decision was made was petition in support of the teacher which was drawn up by the officers of the school's P.T.A. and signed by twenty-five parents.

The parents' committee left the meeting feeling that they had been outmaneuvered by the school administration and the supporters of the teacher among the other parents. They not only lost on the issue they raised but in doing so they apparently

failed to adequately bring their concern for quality to the front rank in the personnel decision concerning this teacher. Furthermore, they were made to appear to be a small group of complainers outnumbered by far by the representative mass of concerned parents associated with normal channels including the P.T.A.

Apparently as a last resort the committee of parents recruited one of their members to run for the school board in the next election. The issues they developed to support their candidate were the issues of board responsiveness to parental concern and a return to the "basics" in education. Documenting the latter issue was the public airing that the achievement scores of the district's pupils were only slightly above the national averages. The candidate declared these scores to be inadequate for an academic community. While the candidate of the Committee of Parents lost the election, she did succeed in raising issues that were embarrassing to the school board administrators.

Ultimately while it was too late to affect their children the Committee of Parents won their case. First many observers, including other parents and teachers, came to agree with their evaluation of the teacher in question. Second, the principal at the neighborhood was relieved by the central administration and transferred to a non-supervisory administrative position. Third, the new principal tried to ensure that the "free time" pattern was corrected and that the stated educational objectives were pursued in the classroom in question. Fourth, when it was apparent that this last point was failing, the decision was made not to continue the teacher in question beyond the second probationary year.

Several general points can be made about this case. It illustrates the interdependence that develops between an administrative agency and its long-standing clientele groups. Parental groups supported the school administration by channeling information from parents at large about the response to schools' program.

For example, critical parents first voiced their feelings at the local P.T.A. before taking a more active step. In addition the same organization was used by the administration to provide supportive documentation at the local level when a decision had to be made about whether to retain the teacher for a second year. Therefore, clientele groups can help to approximate a "feedback" function for agencies when no other mechanism is available for this activity. In addition the close working of clientele and agency personnel helps to develop and maintain a "community of thought" about the proper policy goals of the agency. Parents and staff agreed with the "open classroom" approach to education in the school in question. This matter of educational policy was consistent with the values and beliefs of the active parents. In providing a policy that the active parents preferred, the school could call upon the support of these activists when "feedback" functions were needed. If the P.T.A. became an arm of the school administration, so too the school became an arm of the average parent in so far as educational preferences were concerned. Living in this neighborhood and sending their children to this school were then not conceptually different from providing their children with books or educational television; all were expressions of policy preferences. In short the "modal parent" was not coopted through trickery but supported the school because it did what he or she wanted and the harmonious interrelationship between the school and the P.T.A. was based upon this fact.

This case also illustrates the test of this interdependence when a new group arises with a somewhat different policy preference. In order to maintain the existing symbiotic relationship the administrator may attempt a variety of temporizing strategies to deal with a new demand. The strategy of symbolic reassurance such as that used by the principal with the Committee of Parents apparently is often effective. Another temporizing strategy is made possible by the "multiple points of access" characteristic of American public processes. The Committee of Parents

were perhaps peculiar in that they didn't give up in frustration as they failed to achieve concrete responses at any step from the teacher to the principal, to the P.T.A., to middle level officials, to the School Board, and finally to the superintendent. A final temporizing strategy is to delay the formal recognition of a demand until after it can be implemented according to normal procedures. The Superintendent's hearing of the Committee of Parents may have been such a procedural strategy. In general new proposals can often be delayed and such temporizing can provide information on how serious and widespread the demand is.

The reason that delay was important in this case was that the personnel decision concerning the particular teacher had serious implications for the general policy of the "open classroom." Moreover, this general policy represented an accommodation between the school administrators on one hand and the parental activists on the other. A specific decision that called the policy into question therefore would have the potential of threatening the political balance established by the school as a public agency and the activist parents as the school clientele group. Thus it is possible to account for much of what happened in this case by recognizing it as an example of the clientele-agency relationship that occurs when agencies are organized in particular ways.

This case also enables one to refine the nature of the specific clientele that to which schools generally adapt. First of all parental groups may often place other values over academic achievement when they think about decision priorities in the schools. In part these value preferences are a result of the complex exchange between one's general value structure that results from social experience - which tends to be linked geographically to neighborhoods and communities in the United States - and other experiences such as participation in school decision making. For example in this case a neighborhood had a relatively homogenous value structure due to the socio-economic similarity of the families

living there. While their socio-economic backgrounds would lead one to predict a preference for ranking academic achievement first among school priorities, upon closer inspection we see that the academic backgrounds of these families in fact led them to rank a particular educational policy as process factor first.

Second, outcomes such as ranking other values over achievement, do not always mean that the school bureaucrats get their way in opposition to the wishes of the community clientele and the expression of representative institutions. Instead it may mean that the administrators and the clientele have come to an agreement upon a particular school policy or procedure, such as the "open classroom" and that the relationship between that policy and achievement is not known or is only guessed at. If the "open classroom" did not always lead in a clear-cut way to increased achievement levels then we could have representative institutions bringing about a close clientele-agency link over policy on process or procedures which inadvertently leads to a undesired achievement outcomes. Thus the empirical questions in educational policy are not only whether socio-economic background differences bear any such relationship to achievement differences but also whether socio-economic factors related to differences in school characteristics such as procedures and processes and whether school characteristics in turn relate to achievement differences.

Third, the definition of the clientele of the schools must be the parents in the school's attendance unit rather than the direct consumers of the school's services, the pupils. The variables we are interested in concerning these parents are their socio-economic characteristics. This definition explicitly introduces political, organizational, and social dimensions into the clientele concept. Thus adults have access to the political process whereas children ordinarily do not. Moreover, it is the adult population as it is structured into neighborhoods which is the relevant conceptualization given the decentralized

organization of the public schools. Finally it is the modal socio-economic characteristic of adults so structured that is the relevant variable, given the tie-up between socio-economic background and values on one hand and neighborhood homogeneity of backgrounds in the United States on the other.

Thus we see that it was because, rather than despite the fact that it was situated in an academic neighborhood that the school in the case study placed other values over academic achievement as it rank ordered its priorities. It did so because it thereby adopted a set of classroom procedures most consistent with the value structure of the average academic and persons with similar socio-economic characteristics. Because of that consistency the school could call upon the support of the parents as clientele. The fact that these classroom procedures did not result in increases in academic achievement in one case might be the result of a goal displacement strategy on the part of one teacher-the substitution of "free time" for "open" - rather than an unresponsive public school administration.

In general the case illustrates that there can be a close articulation between the values of parental groupings and the educational policies pursued by the schools. In fact given representative institutions on the one hand and the high degree of decentralization of school decision making on the other hand we should be surprised at any other pattern. Thus the parents of the children in the school have the incentives and the means to influence school policy. One therefore would hypothesize a broad clientele hypothesis of a close articulation between school policy and circumstances and parental values.

While this hypothesis sounds like common sense at first glance, consider its implications for the body of literature about student achievement and the further implication of the relation between values and parental socio-economic status. It is well known that values spring from among other sources, the set of experiences summarized in the sociological measures of status. Values in this sense are merely preference statements that reflect different class positions, etc.

For example the academic parents in our case study made statements about preference for achievement oriented policies or for the "open classroom." Their statements are merely symptoms of the basic structural relation between their background characteristic, academic occupation, and their policy preferences. Thus without knowing their statements we could have predicted as a matter of social science that there would be the structural relation that obtains. There is therefore, a clientele-structure hypothesis that predicts a background-educational policy relation given the status factor and the additional circumstances of decentralization and representation. Why then would one want to "control" or hold constant parental background factors when examining the effects of school differences, as do the scholars in the literature on achievement? The clientele-structure hypotheses says that these factors are closely articulated as the local school strives to conform with the needs and preferences of its clientele, the parents of the children in the school. We would then not expect these factors to have independent effects, they are part of the same process of political influence. If they are not independent controlling one is the same or controlling the other, conceptually speaking. Since a variable when held constant has no measurable relation to another, we would not expect school factors to be related to achievement differences once family characteristics are controlled. Thus, in attempting to separately manipulate different variables that might have causal influences on achievement the previous literature made the mistake of ignoring political processes. These processes would seem to indicate that two of these broad categories of variables are not independent and therefore should not be controlled in the ordinary way. To avoid this mistake we will first specify a set of relations derived from the clientele-structure hypothesis. These relations will then guide the choice of statistical controls and operational definitions of variables.

Procedures:

The Clientele-Structure Hypothesis of Educational Policy

The logic of the argument so far has been that the procedural characteristics of the school serve as intervening variables which "translate" or convert public preferences that are associated with parental background characteristics into performance or "output" factors in the schools, such as measured achievement. Thus in this case study we could not understand how academic parents could support a local school that placed gains in academic achievement in less than first priority until we took into consideration the intervening factor of the "open classroom" approach as a school characteristic. This logic depends upon an integration of the conventional theory of organization in public administration with a social structural theory of political participation. The integration of these theories then leads to a specification of the functional relations which serve as the bases for the empirical tests.

The surface features of the organization of public schools in the United States would lead one to conclude that they in their high degree of decentralization serve as examples of organization by place. However the foregoing discussion indicates that organization on that basis is not inconsistent with organization on other bases as well, despite the conclusions of Herbert Simon. In fact, it would appear that clientele groups may develop around most administrative agencies no matter how they are organized because of the desire of individuals and groups to gain greater influence over decisions that affect them. While citizens in general may have incentives to influence agencies, it would seem to be the case that that sub-set that is most directly affected would have the greatest incentive to gain influence. In similar fashion

agencies may have incentives to foster or to reinforce the establishment of clientele groups in order to reduce uncertainty about public support for programs and policies. With respect to the public schools we often see the parents in the neighborhood attempting to participate in the decisions of the local school. They therefore serve as a clientele group supporting the neighborhood school once it is established and influencing and responding to its programs and policies. Thus the parents in the neighborhood that serves as a school's attendance unit act as members of the local unit which influences school programs through the normal pattern associated with the politics of "organization-by-place" and they also serve as the basis of the clientele of the local school as if "organization-by-clientele group" also occurred. We then have in effect two different organizational patterns superimposed upon each other, both of which can be predicted to increase neighborhood articulation with the program of the local school. We therefore have very powerful forces moving toward this articulation. In general then we would expect a functional relationship in which differences between schools, to the extent such differences can be observed, would vary as a function of neighborhood characteristics, everything else being equal. Thus

$$S = f(N, U_s) \quad 1.)$$

where S = school characteristics
 N = neighborhood characteristics
 U_s = influences on S that are
independent of N .

Therefore this equation specifies a functional articulation between the nature of the neighborhood and the neighborhood school. As such it serves as a very general hypothesis of the study. The fact that everything else isn't always equal and therefore we can expect some "slippage" between school and neighborhood

characteristics is explicitly allowed for in this hypothesis through the factor, U_s .

Now the reason that clientele groups attempt to influence and support agencies is that they expect that the programs of those agencies will have some effect on things they care about. This is what we usually mean by performance, impact, or output. Thus groups support the "open classroom" approach because they think it will lead their children to learn self-reliance, toleration, democracy, and will allow them to develop academically at their own "pace." In supporting programs clientele groups undoubtedly are operating in terms of a hypothesis in which a performance characteristic of some sort is predicted to be dependent upon the program characteristics of the school. While groups might make mistakes about such hypotheses - indeed educational professionals would often be hard-pressed to provide guidance in the form of research and theory on these matters - we might expect some levels of perceptual accuracy to develop over time and yet still allow for "slippage" in the hypothesis. Therefore if clientele groups are at all accurate we would find performance indicators to be a function of school characteristics. Thus

$$P = f(S, U_p) \quad 2.)$$

where P = school performance indicators
 S = school characteristics
 U_p = influences on P that are independent of S .

Equations 1.) and 2.) are general hypotheses specified by initial observations of the relations between parents and schools given the way the public schools are organized administratively in the United States. In order to operationalize these hypotheses additional considerations will have to be introduced by way of further specification. Before doing that however a third general hypothesis can be specified. This third hypothesis is a derivation

from the first two. As a derivation it is implied by the theory under consideration but is not directly dependent upon it. Very simply substituting equation 1.) into equation 2.) gives third functional relation. Thus

$$P = f[f(N, U_s), U_p] \quad 3.)$$

where all factors are defined in 1.) and 2.) above.

Equation 3.) shows that neighborhood characteristics can have a causal effect on school performance because of, not despite, the effect of the schools on performance measures such as achievement tests. This causal effect will show up in empirical analyses where neighborhood socio-economic characteristics or their equivalents are the primary independent variables and achievement tests are the dependent variables. The prediction of this effect is the primary hypothesis of this study and it is derived from the theory of the political organization of the public schools. The test of this hypothesis is not straight-forward but instead requires a somewhat involved statistical procedure. In order to introduce that procedure it will be instructive to consider a contrasting, less political view of the place of socio-economic variables in achievement.

Most of the previous research on achievement has treated socio-economic family background of the child as a "confounding" factor that must be held constant when one examines the effect of school-related practices on achievement differences. This research strategy implicitly assumes that school and socio-economic factors have independent influences on achievement. Statistical control procedures are then invoked, holding constant the two types of factors in order to estimate their independent effects.

The error in this strategy is apparent when the implications of equation 3.) are hypothesized. If equation 3.) is correct then the family background

measures in large part are simply surrogates of the socio-economic character of the neighborhood. These surrogate variables are measured at the level of the individual child rather than at the more aggregative neighborhood. However they refer to the same feature when the achievement scores of children from different schools are compared. This feature is the different socio-economic status tendency of the children in the school and consequently of the school itself. This status tendency is not independent of the educational practices of the school and can not be treated as such if the hypothesis is correct. Instead status differences result in differences in school practices through complex political relationships. Controlling for status differences therefore would be the functional equivalent for controlling for differences in school characteristics and vice-versa. It is this interdependence versus the previous literature's implicit assumption of independence between school and family characteristics that is the key to the logic behind the empirical test of the hypothesis.

In order to further specify the hypotheses associated with equations 1.)-3.), we must consider several issues. The first order of priority is to deal with the issue of the measurement of the variables. Because the hypotheses represent a major change in the interpretation of the roles of school and family factors as compared to the interpretations offered in the past literature, it was decided to use the data base and measures of these variables upon which much of the past literature depends: the "Coleman report data." Thus the research strategy includes the notion that the "Coleman data" offers the best support for the Coleman interpretation (and related interpretations). Therefore a new interpretation would have its severest test with these data. Readers interested in the precise form of the questions that underlie the "Coleman data" should refer to the original report of that study. The general descriptions

of the variables actually used from that study are listed in Appendix B.

The next issue is the form of the variables to be used in the tests. Since the hypotheses refer to school-by-school comparisons it was decided to use the data in the form of variables "aggregated" to the school level. Besides being appropriate to the hypotheses this choice has the effect of reducing the random elements captured in the individual level of the data, thereby increasing the value of R^2 and reducing this source of bias in the estimates of the regression coefficient. On the other hand it may confound non-random measurement "error" with the effects of the independent variables and thereby lend to "spurious" results. Each of these possible outcomes is addressed below.

First of all the possibility of random measurement error has not been taken seriously in the literature. In fact one of the interpretations of the low R^2 's obtained with individual level data is that "luck" is the primary explanatory factor influencing achievement. In order to make this inference we would have to assume perfect ability to measure achievement. Given the present "state of the art" of pencil and paper achievement tests this assumption would appear wholly unwarranted. Appendix C gives the implications of random measurement error in all of our variables for complex policy models of educational achievement.

The primary implications of concern are that with error in the independent variables we would have a "downward" bias in the estimated values of regression coefficients and that if there were differential error across variables we would have differential failure of the ordinary tests of statistical significance. For example if family background factors can be measured with more precision than can "school characteristics" and if the variables are related as equation 2.) specifies then an attempt to compare the "independent"

effects of both on achievement would more often lead to the conclusion that school characteristics are statistically "insignificant" than family background factors. This conclusion however would be due to a combination of differential measurement error and failure to take equation 2.) into account. The tests of significance employed in this study will avoid this difficulty by employing aggregative measures of the variables and by proper selection of analytic strategies based upon the three specified equations above.

Second of all it is possible that measurement error may not be random but it may be systematic due to the "causal" variables that have been left out of the analysis. Under these conditions the systematic elements could be confounded with the effects of the measured variables as they are aggregated. Such confounding influences would appear as part of the estimated effects of measured variables when simple bivariate relationships are examined. In short the aggregation of variables could produce spurious associations due to the systematic effects of "left out" variables.

When we conceptualize the aggregation problem as the spuriousness problem we immediately are led to two conclusions. On the one hand there is a sense in which a given analysis never solves this problem in that because of the lack of theory we can not anticipate all unmeasured influences on a given process or relationship. The ultimate solutions are a matter of the cumulative development of knowledge by the scholarly community over a period of time. On the other hand short run attempts to solve the problem do exist. These are the attempts at measuring and bringing into the analysis those factors that might be hypothesized as producing spuriousness given the current state of knowledge.

From this perspective we can interpret the research strategies in the previous literature as hypothesizing that parental background would serve as such spuriousness-producing factors when school variables were related to

achievement scores. The implicit or explicit theory seems to be that some educational processes take place outside of the school, especially in the case of middle class and upper middle class families. If we don't first account for these processes we might confuse their effects with that of the school. Similarly if we aggregate to the neighborhood school level, we might inadvertently confound these influences with the school variables.

In order to guard against such spuriousness family background factors similarly aggregated will be explicitly introduced. However the "controls" will not be automatic but will be systematically introduced according to the logic specified by equations 1.)-3.). In addition a second set of "control" variables will be introduced following the logic of this outside-the-school learning theory. Thus some of the Coleman report's measures of the family situation include factors that appear to describe attempts to provide learning situations in the home. If we assume, along with this theory, that these "other learning" attempts are related to the social background characteristics of the neighborhood, then an additional set of equations must be entertained. Thus

$$L = f(N, U_L) \quad 4.)$$

where L = non-school learning factors
 N = neighborhood characteristics
 U_L = influences on L that are
independent of N .

$$P = f(L, U'_p) \quad 5.)$$

where factors are as before

and by deduction

$$P = f[f(N, U_L), U'_p] \quad 6.)$$

where factors are as before.

Comparing equation 6.) with equation 3.) we see that there are two processes by which we can deduce a relationship between the socio-economic characteristics of the school neighborhood and its performance measured by factors like student achievement. One is that middle and upper-middle class parents provide educational situations for their children in the home. The other is that middle and upper-middle class parents provide educational situations for their children in the neighborhood school. The previous research emphasized the former process. The present research adds a consideration of the latter process by showing the importance of the political and administrative organization of the schools. The possibility that both processes take place then call for the use of appropriate control procedures in analyzing the effects of each.

The final steps in specifying the appropriate models call for consideration of complexities involving non-linear and non-additive relationships among these variables. While there are several different theoretical possibilities of such patterns, the empirical examinations carried out in this research failed to demonstrate support for this area of complexity. A total of sixty different scatter-plots where non-linear effects might have been expected failed to show anything but closely linear patterns. Similarly data generated by the research group associated with Christopher Jencks were kindly made available to the present study in order to examine theoretically specified non-additive effects. A persistent lack of significant non-additive effects were observed. Consequently the complex models reported here are linear and additive.

The complexity of the study then is that associated with the theoretical specification of collinearity between the independent variables in this field and of measurement error in these variables. The procedures designed to deal

with these problems involve then a form of linear regression in which instrumental variables are applied to a problem of multicollinearity.

In summary then the procedure necessary to test the more complex models that have been deduced would be to utilize linear regression equations in which achievement scores are dependent upon different combinations of independent variables. The following section gives the equations used and the results. The results were reported in terms of a procedure that looks at the coefficient associated with an independent variable in a regression equation as summarizing the influence of that variable upon the dependent variable.

Results: Tests of Hypotheses

The tests of the basic hypotheses rest upon a regression equation of the form

$$y = a + b_1x_1 + b_2x_2 + c z + e$$

where y = achievement scores

x_1 = school characteristics

x_2 = "other learning" reported factors

z = social background characteristic

e = error term

a, b_1, b_2, c = regression coefficients and all

variables are aggregated to the

level of the school attendance

area.

In the various tests the significance level will be based upon the conventional rule of thumb which requires the coefficient to be approximately twice the standard error of the estimated value. Similarly the test of a change in a coefficient would have to have the same ratio. Because each variable in the above equation refers to a whole host of "operational" indicators and because these indicators are themselves correlated within and between variable categories the research strategy is more complex than simply estimating the above model. The greatest complexity depends upon the interrelationships of the independent variables.

Of course the fact that school characteristics are correlated with neighborhood and family background factors is well known. Moynihan and Mosteller even point to this association as the source of the problem of collinearity in past research on achievement. What we are attempting in the present study is

not to "reinvent the wheel" by pretending to discover the school-family relationship. Instead we are trying to show the importance of this pattern throughout the entire family-school-achievement process. Consequently we have chosen an analytic strategy which does not calculate the correlation between the raw scores that measure family background and school characteristics. Instead an "instrumental variables" (See Appendix D) strategy has been adopted. In this strategy we first obtain the predictive or systematic elements of these two "independent" variables for achievement by separately purging them of random elements in terms of their relation to the dependent variable, achievement scores. This step has the effect of removing the random elements of U_s and U_p in equations 1) - 3). We then examine the interrelation of the two purged versions of the independent variables. The resulting interrelationship shows us the association between the two variables that is important for their separate abilities to predict the dependent variable. Thus if they were truly independent in their affects on achievement the two purged or instrumental variables would be uncorrelated. On the other hand the extent of their correlation is a part of the measure of their participation in the entire process that proceeds from the average family background characteristic in the school area to the nature of the school itself and finally to achievement levels of students in the shcool.

What is the nature of the relationship between these predictor variables? Table 1 gives the values for the square of the correlation coefficient, the slope coefficient of the regression equation, and the standard error for this relationship in terms of each achievement variable measured for each of the three grades. (See Appendix A for the entire set of estimates.) In all cases the relationship stands up under any conventional statistical significance test.

Table 1: Regression Estimates from the Equation,
 School Instrument = a + b Family Background Instrument,
 by, Dependent Variable and Grade

<u>Dependent Variable</u>		<u>6th Grade</u>	<u>9th Grade</u>	<u>12th Grade</u>
Non Verbal Score	Coef	.1737	.9448	1.6613
	SE	(.0034)	(.016)	(.030)
	R ²	.5210	.7901	.7962
Verbal Scale Score	Coef	.1790	.5461	.7447
	SE	(.0029)	(.009)	(.013)
	R ²	.6156	.8164	.8113
Reading Comprehension	Coef	.2908	.8203	1.5809
	SE	(.0040)	(.017)	(.027)
	R ²	.6821	.7057	.8203
Math Achievement	Coef	.2808	.5427	.7636
	SE	(.0040)	(.010)	(.015)
	R ²	.6727	.7447	.7759
General Information	Coef		.5570	.9031
	SE		(.009)	(.016)
	R ²		.8015	.8137

Moreover the square of the correlation coefficients are surprisingly large values. They are all larger than .5 which means that more than fifty per cent of the variance is explained in every case. Moreover in twelve of the fourteen relationships the level of explanatory power reaches from about the seventy per cent level to higher. In half of the cases the proportion explained is in the eighty per cent range. Overall the school-family predictor variables show a variance explained that averages just short of seventy-five percent, a remarkably high level of predictive power in the social sciences.

The interpretation of this rather close relationship is that school characteristics are dependent upon the background tendencies of the families whose children form the school attendance area. Thus the regression equations which serve as the basis for the calculation of the proportions of explained variance treat school characteristics as the variable to be explained and family background as the explanatory variable. The resulting regression coefficient estimates shows that a unit change in the family background instrument produces about a .7 change in school characteristics on the average with a range of latter values from .17 to over 1.6. This interpretation supports the conclusion that the reason that schools vary in their internal characteristics-- to the extent that they do vary-- is that the backgrounds of the families of the students from their attendance areas vary at least in so far as these factors are predictive of student achievement.

The second phrase in spelling out the role of school factors in a hypothetical family-school-achievement sequence is the examination of what happens to the family-achievement relation when the school influence is factored out. We know of course from past research that family background is highly influential in any measure of achievement differences. Such influences

undoubtedly stand for a whole set of status differences including social class and racial and ethnic patterns. Some researchers and even some school officials read this research as measuring that the family has the only influence on achievement differences and that the school plays no role.

In order to further examine this hypothesis we must attempt to gain a measure of what would occur if the school differences that do obtain empirically were somehow set aside. The argument behind this logic is that the patterns that can be studied with respect to the schools have already occurred--they are not subject to the experimental manipulation of the researcher. Therefore we must settle for the substitution of statistical for experimental control as the basis for inferring causal affects. Furthermore our theory and past research gives certain important clues as to the nature of these patterns. First the organization of American public schools produces a strong relationship between family and school differences. Secondly, we know that there is another strong relationship between family and achievement differences. Therefore we are left with two possibilities: Achievement differences are independent of school differences in a pattern in which both rest upon family background as part of independent, non-causally related processes. Achievement differences are dependent upon school differences in a pattern that begins with family influences on schools and school influences upon achievement. An inference that the latter process is appropriate would lead to the conclusion that schools play a major role in bringing about the family-achievement relationship while the former would lead to the conclusion that this relationship occurs independently of school differences. Therefore a necessary test for these alternative differences would be to examine the family-achievement relationship before and after school differences are factored out of the equation. If this relationship is relatively

independent of school differences we would expect no statistically significant differences in it before and after school differences are "removed" from the equation. On the other hand statistically significant differences would lead to the inference of a family-achievement process interpreted or brought about through school differences.

The procedure developed for the statistical test depends upon logical deductions from these alternative patterns. If the school factors do play a role in the achievement process and strong relationships exist between family and school on one hand and family and achievement on the other then an estimate from a regression equation that includes only achievement and family variables will be biased. In fact if the variables are all treated as positive influences and are as stated the direction of the bias can be predicted; there will be an upward bias in the achievement-school estimate because the estimated family influence will mask or include the school influence as well. On the other hand when school variables are explicitly included with family variables in the equation predicting achievement this bias in the family estimates will tend to be factored out. We therefore expect a reduction in the family estimates when school variables are added. However if achievement is independent of school factors and truly rests only on family differences then the family estimates will tend to be free of this source of bias. Consequently we would not expect a statistically significant reduction in the estimated values of the family influence when school variables are entered into the regression equation. Therefore the test consists of comparing the values of the family background estimates before and after school variables are entered into an equation in which achievement scores are the dependent variables. If these values show statistically significant differences we reject the null hypothesis that they are the same--that the family did not mask school influences in the "before"

estimates--and infer the family-school-achievement causal pattern.

The instrumental variables approach to this procedure in fact produces the "before" and "after" estimates in reverse order. We derive a set of independent variables and estimate their influence in two stages.

In the first stage all independent variables are entered into an equation to predict achievement scores. From this equation we obtain our instrumental versions of the school and family variables. These are the portions of the school and family measures that predict achievement when both types of independent variables appear in a given equation. These instrumental variables are then utilized in a second stage which is simply a second regression equation predicting achievement from the instrumental versions of both variables. Since the instruments have been calculated in terms of their predictability their "slope" coefficients from the second stage equation have a convenient property: they have the value of 1.0 except for minor errors of calculation.

These second stage estimates are those which represent the separate factoring out of both family and school upon the dependent variable. These estimates must be compared with another set where family factors alone appear in the equation, i.e., where school influences are not factored out. To obtain these we utilize yet another regression equation where the family background instruments alone serve as the independent variable. If the family factors mask the school factors this last set of estimates will demonstrate an upward bias in comparison with family estimates from the equation with both family and school instruments. We therefore can make use of the convenient value of the coefficient from the two instrument equations to make this test: We merely take the difference from 1.0 of the coefficient of the family instrument from the equation where that factor is the sole independent variable. The ratio of that difference to the standard error of the standard error of the family coefficient then is

the basis for the test of significance for the differences of the coefficient before and after school differences are "controlled."

Table 2 summarizes the results of these complex analyses. The table gives the ratio (t) of the differences of the coefficient of the family instrument from 1.0 to the standard error of that coefficient when the family instrument only appears in the equation. When we consider that the conventional value of approximately 2.0 for this ratio as necessary for rejecting the null hypothesis of no differences we see that the results are both striking and consistent. We achieve conventional standards of statistical significance by far and we do so in every case.

The ratio shows some interesting variability. In general the values increase with age so that they are larger for twelfth grade and smallest for the sixth grade. Within the sixth grade the ratio is largest for mathematical achievement and smallest for non-verbal scale scores. These patterns are consistent with measurement error hypotheses that expect smaller random behavior captured in the dependent variables as age and precision of subject matter increases.

A similar test can be carried out with respect to another set of family characteristics. In addition to family background factors the Coleman study obtained measures relating to family behavior that correspond to learning what takes place in the home rather than at the school. Two questions can be asked of these variables: Do they also appear as part of the family-school-achievement sequence and therefore can be interpreted with the school variables? Do they also serve as interpreting factors themselves, standing between family background and achievement.

Table 2
 Test of Significance of Family Background Instrument Coefficient
 with and without School Instrument in the Equation

	<u>Sixth Grade</u>	<u>Ninth Grade</u>	<u>Twelfth Grade</u>
Non Verbal Scale Score	14.4 (.173/.012)	36.3 (.945/.026)	40.5 (1.661/.041)
Verbal Scale Score	17.9 (.179/.010)	42.0 (.546/.013)	39.2 (.745/.019)
Reading Comprehension	22.4 (.291/.013)	34.2 (.820/.024)	45.2 (1.581/.035)
Mathematical Achievement	23.4 (.281/.012)	30.2 (.543/.018)	31.8 (.764/.024)
General Information		37.1 (.557/.015)	39.3 (.903/.023)

Table 3 gives the ratios to test whether school variables play a role in interpreting these additional learning variables. Once again the ratios for surpass any conventional level imposed by ordinary significance tests. The inference that can be made then is that practices designed to foster learning in the home and type of school serve as part of the same overall pattern of family choices that result in measured achievement differences.

Finally Table 4 gives the ratios based upon the differences in family background coefficients when both school and family learning practices enter the equation. As before the ratios amply demonstrate significant differences in estimates with and without the additional independent variables in the equation.

The inference that must be drawn from these tests is that the family background differences that ordinarily conform to American status distinctions do not result in differential achievement through some sort of magic. Instead these differences are a part of an overall process in which social status results in predispositions toward learning that not only increases learning in the home but also in school differences that bring achievement differentials. As a result the school plays an important role in bringing about achievement differences. This role is one set firmly in the context of the differences in American family backgrounds given the way we have organized the American school system. One must conclude then that achievement differences are due to school differences because in large part, the reason that family characteristics correlate with achievement differences is because of the differences in schools that those families send their children, given the neighborhood where they live.

Table 3
 Test of Significance of Family Learning Instrument
 With and Without School Instrument in the Equation

	Sixth Grade	Ninth Grade	Twelfth Grade
Non Verbal Scale Score	40.9 $\frac{(2.35-1.00)}{.033}$	23.6 $\frac{(4.40-1.00)}{.140}$	36.7 $\frac{(3.97-1.00)}{.081}$
Verbal Scale Score	53.9 $\frac{(2.67-1.00)}{.031}$	36.4 $\frac{(5.84-1.00)}{.133}$	28.0 $\frac{(5.57-1.00)}{.163}$
Reading Comprehension	46.2 $\frac{(2.71-1.00)}{.037}$	23.4 $\frac{(4.37-1.00)}{.144}$	29.9 $\frac{(8.23-1.00)}{.242}$
Mathematical Achievement	28.9 $\frac{(3.27-1.00)}{.044}$	15.3 $\frac{(5.29-1.00)}{.28}$	26.9 $\frac{(5.12-1.00)}{.153}$
General Information		35.9 $\frac{(4.91-1.00)}{.109}$	24.0 $\frac{(5.96-1.00)}{.207}$

Table 4
 Test of Significance of Family Background Instrument
 With and Without School and Family Learning Instruments in
 the Equation.

	<u>Sixth Grade</u>	<u>Ninth Grade</u>	<u>Twelfth Grade</u>
Non Verbal Scale Score	32.2 (1.58-1.00) .018	39.0 (2.21-1.00) .031	(3.14-1.00) .050
Verbal Scale Score	43.6 (1.61-1.00) .014	46.9 (1.75-1.00) .016	37.0 (1.74-1.00) .020
Reading Comprehension	39.4 (1.67-1.00) .017	34.8 (1.94-1.00) .027	45.9 (2.47-1.00) .032
Mathematical Achievement	45.8 (1.87-1.00) .019	32.0 (1.64-1.00) .020	32.9 (1.79-1.00) .024
General Information		46.3 (1.74-1.00) .016	35.9 (1.79-1.00) .022

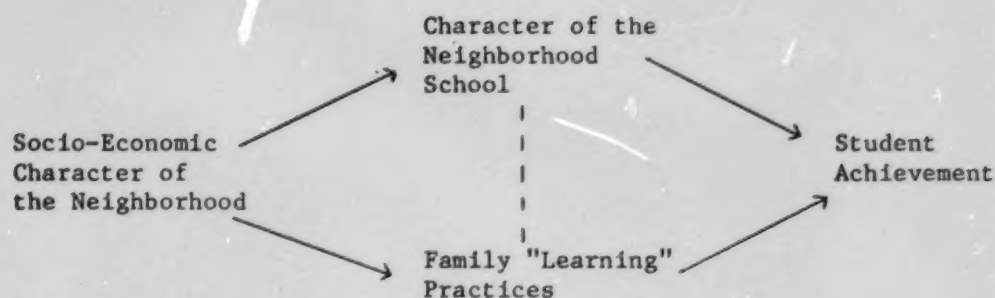
Conclusions

This analysis has depended upon a total of nearly forty regression equations constructed in sequence according to theoretical criteria and involving over two hundred separate variables. The appendices give the entire set of estimates, first for the equations involving the instrumental variables (Appendix A) and then for the "raw" variables used for the construction of the instruments. In addition, given the size of the data set and the high degree of colinearity between the variables of a given theoretical category, a great deal of preliminary work has gone into "boiling" down the number of independent variables to a manageable subset. These "deletion" runs were regression equations in which all variables of a given category, like family background variables were used to predict the dependent variable. Any variable that did not demonstrate a significant effect at the .05 level was deleted thus reducing the level of interrelationship within that category of variables so that the role of interrelationship across categories (i.e. the interrelationship between family background and school characteristic) would be allowed full play. This process was repeated for each set of independent variables. Thus a great deal of preparatory effort went into bringing the data into a form that the theory in question could be examined.

The examination indicates that the previous research misinterpreted the role of social economic factors in achievement differences. In large part these factors exist as part of the process by which school differences relate to achievement differences, rather than as independent of that process. As a result the impact of school factors were underestimated in that research.

This research shows that school differences go hand in hand with achievement differences even when social economic factors are taken into account in a theoretically proper manner. As a result it appears proper to infer that schools do make a difference in achievement. At least one can conclude that school differences have been prematurely dismissed as playing a role here.

In terms of a "path" model of the sort that has become popular in educational research, one can conclude that the process that operates in our schools could be illustrated as follows:



That is, socio-economic and other measures of the family situation are highly related to the character of the school the family's children go to due to our political decisions on the "neighborhood" organization of schools. As a result those family background factors relate to school outcome or performance variables like achievement, because of, rather than independent of, the nature of the school. This analysis therefore lead to a new emphasis on the importance of the school in achievement differences; and therefore it leads to a call for a return to research on what aspects of school practices lead to higher achievement results.

Recommendations

Research on a less aggregative level is necessary to pin down just what it is in the high social economic status neighborhoods that lead to higher achievement scores. The larger inference from this study is that those practices could be transferred to other schools with appropriate investment, training, and perhaps oversight of the school. At least the trend of moving away from the consideration of question of this sort over the past decade can be shown to rest on a faulty premise. That was that the schools didn't make any difference anyway. Therefore we didn't need to know what went on in them and they couldn't be used effectively to equalize achievement and eventually job success. The analysis here shows those conclusions can not be based upon the evidence they claimed to depend upon. The reason for this difference in conclusion was that the earlier research did not consider an adequate model of the political process within which our schools operate. Consequently it made erroneous statistical assumptions and thus erroneous inferences.

Achievement differences in the aggregate across American schools are not independent of school differences nor are school differences independent of differences in the family characteristics in these schools. From the perspective of a conception of the school as a public agency we should not be surprised by these facts. The schools are organized in a way to accentuate neighborhood by neighborhood differences in their practice and policies, despite the nationalizing tendencies of professionalism and mass communication. Some of these practices and policies have to do with different degrees of emphasis being placed on administrative goals or outcomes such as achievement. Since emphasis on achievement is one of the ideas of a middle to upper-middle socio-economic

status orientation, it is not surprising that such family background characteristics go along with school differences and these differences with achievement differences. The real surprise is in how this process was ignored in the previous research and how that research is now being used to remove the schools from the discussion and research on agencies available to create greater social equality.

Bibliography

1. Aldrich, John, and Cnudde, Charles F. 1975. "Probing the Bounds of Conventional Wisdom: A Comparison of Regression, Probit and Discriminant Analysis" American Journal of Political Science (August, 1975): 571-608
2. Armor, David. 1972. "The Evidence on Busing," The Public Interest (Summer 1972): 90-126.
3. _____. 1972-A. "School and Family Effects on Black and White Achievement: A Reexamination of USOE Data" in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 168-229.
4. Blalock, Hubert M., Jr. 1969. Theory Construction (1969).
5. Cain, Glen G. and Watts, Harold W. 1970. "Problems in Making Policy Inferences from the Coleman Report," American Sociological Review (April 1970): 228-242.
6. Cnudde, Charles F. and McCrone, Donald J. 1969. "Party Competition and Welfare Policies in the American States," American Political Science Review (September 1969): 858-866.
7. Cnudde, Charles F. 1972. "Theories of Political Development and the Assumptions of Statistical Models: An Evaluation of Two Models," Comparative Political Studies (July 1972): 131-150.
8. _____. 1972-A. "A Specification of State and National Differences in Electoral Behavior," American Political Science Association Convention (September 1972).
9. _____. 1973. "Stochastic Processes and Political Theory," Comparative Political Studies (July 1973).
10. _____. 1973. "Social Status and the 1968 Election in California," in James W. Prothro, et al., eds. Explaining the 1968 Election (1973).
11. _____. 1974. "Complex Policy Models of Educational Achievement: Multicollinearity in the Sixth Grade." Conference on Social Science and Education, (May 1974).
12. Coleman, James S., et al. 1966. Equality of Educational Opportunity (1966).
13. Coleman, James S. 1966-A. "Equal Schools or Equal Students," The Public Interest (Summer 1966).
14. _____. 1968. "The Mathematical Study of Change," in Hubert M. Blalock and Ann B. Blalock, eds., Methodology in Social Research (1968): 428-478.

15. _____. 1970. "Reply to Cain and Watts," American Sociological Review (April 1970): 242-249.
16. _____. 1972. "The Evaluation of 'Equality of Educational Opportunity'" in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 146-167.
17. _____. 1976. "Response to Professors Pettigrew and Green" Harvard Educational Review (May, 1976): 217-224.
18. Hanushek, Eric A., and Kain, John F. 1972. "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy" in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 116-145.
19. Jencks, Christopher, et al. 1972. Inequality: A Reassessment of the Effects of Family and Schooling in America (1972).
20. Jencks, Christopher. 1972-A. "The Coleman Report and the Conventional Wisdom" in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 69-115.
21. Johnston, J. 1963. Econometric Methods. (1963).
22. Miller, Van, Madden, George R., and Kincheloe, James B. 1972. The Public Administration of American School Systems. (1972).
23. Mosteller, Frederick, and Moynihan, Daniel P. 1972. "A Pathbreaking Report" in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 3-66.
24. Ostrom, Vincent. 1973. The Intellectual Crisis in American Public Administration (1973).
25. Pettigrew, Thomas F., and Green, Robert L. 1976. "School Desegregation in Large Cities: A Critique of the Coleman 'White Flight' Thesis" Harvard Educational Review (February, 1976): 1-53.
26. Rao, Potluri, and Miller, Roger LeRoy. 1971. Applied Econometrics (1971).
27. Simon, Herbert A. 1946. "The Proverbs of Administration" Public Administration Review (Winter, 1946): 53-67.
28. Smith, Marshall S. 1972. "'Equality of Educational Opportunity': The Basic Findings Reconsidered," in Frederick Mosteller and Daniel P. Moynihan, eds., On Equality of Educational Opportunity (1972): 230-342.

Appendices

The following appendices report in more detail the procedures upon which the summary in the body depends. Appendix A gives the full set of estimates underlying Tables 1-4 in the body. Appendix B gives the full set of independent variables utilized in the separate regression equations and the associated estimates. Appendix C gives a proof of the effects of differential measurement error. Finally, Appendix D gives the underpinning of the instrumental variables approach.

Most of these appendices are further supported by additional computer print out, available upon request through the Center for Public Policy Analysis, Michigan State University. These include the stages of the two-stage least squares runs, the full set of "deletion" runs and the tests for non-linearity. These latter materials represent over three-years labor of many many individuals and a mountain of print-out.

Appendix A

Estimates of Relations Involving Instrumental Variables

MULTIPLE AND BI-VARIATE ESTIMATES OF THE ROLES OF FAMILY, SCHOOL, AND ADDITIONAL LEARNING VARIABLES ON ACHIEVEMENT: INSTRUMENTAL VARIABLES

Achievement Variables	Instrumental Variables	Sixth Grade												Ninth Grade												Twelfth Grade											
		Two Variables						Three Variables						Two Variables						Three Variables						Two Variables						Three Variables					
		b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²	b	R ²	c	R ²				
Non Verbal	Family: Z	Coef	1.0014	.8026	1.1740	.7870	.9894	.8239	1.578	.7554	Coef	1.0000	.9129	1.9448	.8590	1.0000	.9213	2.2083	.8465	Coef	1.0000	.9314	2.6613	.8473	1.0000	.9340	3.1388	.8342	1.0000	.9340	3.1388	.8342					
	SE	(.0170)		(.0121)	(.0246)		(.0183)		(.0183)	SE	(.045)		(.026)		(.047)		(.031)		(.031)	SE	(.040)		(.041)		(.048)		(.050)		(.050)		(.050)		(.050)				
	School: X ₁	Coef	.9911		3.9965	.5283	1.0817		5.2618	.2976	Coef	1.0000		1.8362	.8652	1.0000		2.0242	.8515	Coef	1.0000		1.4793	.9074	1.0000		1.8412	.8862	1.0000		1.8412	.8862					
	SE	(.071)		(.0764)		(.0955)		(.1614)		(.1614)	SE	(.042)		(.024)		(.042)		(.028)		(.028)	SE	(.032)		(.017)		(.038)		(.024)		(.038)		(.024)		(.038)			
Add. Learning: X ₂	Coef					.9971		2.3497	.6801					1.0202		4.4018	.5166			Coef					1.0000		3.9703	.7553			1.0000		3.9703	.7553			
	SE					(.0387)		(.0326)						(.075)		(.140)				SE					(.075)		(.081)			(.075)		(.081)		(.075)			
	Verbal	Family: Z	Coef	1.0000	.8409	1.1790	.8487	1.0000	.8798	1.6128	.8357	Coef	1.0000	.9516	1.5460	.9311	1.0000	.9409	1.7534	.9283	Coef	1.0000	.9523	1.7447	.9136	1.0000	.9563	1.7402	.9003	1.0000	.9563	1.7402	.9003				
	SE	(.0155)		(.0103)	(.024)		(.014)		(.014)		(.014)	SE	(.026)		(.013)		(.030)		(.016)		SE	(.033)		(.019)		(.027)		(.020)		(.020)		(.020)		(.020)			
School: X ₁	Coef	1.0000		4.399	.6262	1.0000		6.1099	.4788	Coef	1.0000		2.4950	.8857	1.0000		2.7617	.8707	Coef	1.0000		2.0894	.8937	1.0000		2.5256	.8324	1.0000		2.5256	.8324						
	SE	(.0482)		(.0493)		(.084)		(.129)		(.129)	SE	(.042)		(.030)		(.044)		(.035)		SE	(.040)		(.026)		(.039)		(.041)		(.039)		(.041)		(.039)				
	Add. Learning: X ₂	Coef					1.0000		2.6717	.7568					1.0000		5.8411	.6771			Coef					1.0000		3.0573	.3532			1.0000		3.0573	.3532		
	SE					(.040)		(.031)		(.031)	SE				(.076)		(.133)		(.133)		SE					(.070)		(.153)			(.070)		(.153)		(.070)		
Reading Comprehension	Family: Z	Coef	1.000	.8263	1.2908	.8073	.9961	.8524	1.6910	.7997	Coef	1.0000	.9331	1.8204	.8602	1.0000	.9395	1.9618	.8529	Coef	1.0000	.9464	2.5809	.8746	1.0000	.9486	2.4702	.8827	1.0000	.9486	2.4702	.8827					
	SE	(.0215)		(.0127)	(.0283)		(.0171)		(.0171)		(.0171)	SE	(.031)		(.024)		(.032)		(.027)		SE	(.034)		(.035)		(.051)		(.032)		(.032)		(.032)		(.032)			
	School: X ₁	Coef	1.000		3.3459	.6723	.9887		4.7960	.6136	Coef	1.0000		1.8602	.8347	1.0000		2.1781	.8418	Coef	1.0000		1.5189	.9328	1.0000		1.7660	.9140	1.0000		1.7660	.9140					
	SE	(.0410)		(.0472)		(.0776)		(.0769)		(.0769)	SE	(.032)		(.025)		(.036)		(.031)		(.031)	SE	(.031)		(.016)		(.018)		(.019)		(.019)		(.019)		(.019)			
Add. Learning: X ₂	Coef					.9971		2.7040	.6916					1.0000		4.3473	.4987			Coef					1.0000		8.2282	.5989			1.0000		8.2282	.5989			
	SE					(.0416)		(.0365)		(.0365)	SE				(.065)		(.144)		(.144)	SE					(.112)		(.242)			(.112)		(.242)		(.112)			
	Math Achievement	Family: Z	Coef	1.000	.8332	1.2808	.8141	1.000	.8508	1.8730	.7927	Coef	1.0000	.9245	1.5427	.8864	1.0000	.9281	1.6411	.8837	Coef	1.0000	.9246	1.7636	.8771	1.0000	.9303	1.7852	.8784	1.0000	.9303	1.7852	.8784				
	SE	(.0205)		(.0124)	(.0327)		(.0194)		(.0194)		(.0194)	SE	(.029)		(.018)		(.030)		(.020)		SE	(.039)		(.024)		(.038)		(.024)		(.024)		(.024)		(.024)			
School: X ₁	Coef	1.000		3.3956	.6708	1.000		3.2645	.7046	Coef	1.0000		2.3722	.8289	1.0000		2.4521	.8287	Coef	1.0000		2.0160	.8616	1.0000		2.4614	.7947	1.0000		2.4614	.7947						
	SE	(.0598)		(.0481)		(.0534)		(.0428)		(.0428)	SE	(.047)		(.036)		(.048)		(.037)		SE	(.045)		(.029)		(.047)		(.045)		(.045)		(.045)		(.045)				
	Add. Learning: X ₂	Coef					1.000		3.2737	.6936					1.0000		3.2886	.2734			Coef					1.0000		5.1231	.5912			1.0000		5.1231	.5912		
	SE					(.0543)		(.0440)		(.0440)	SE				(.111)		(.286)		(.286)	SE					(.095)		(.153)			(.095)		(.153)		(.095)			
General Information	Family: Z	Coef									Coef	1.0000	.9516	1.5370	.9224	1.0000	.9556	1.7394	.9246	Coef	1.0000	.9441	1.9031	.8978	1.0000	.9474	1.7905	.8948	1.0000	.9474	1.7905	.8948					
	SE										SE	(.026)		(.015)		(.026)		(.016)		SE	(.040)		(.023)		(.032)		(.022)		(.022)		(.022)		(.022)				
	School: X ₁	Coef									Coef	1.0000		2.4390	.8761	1.0000		2.9474	.8175	Coef	1.0000		1.9801	.8979	1.0000		2.3439	.8441	1.0000		2.3439	.8441					
	SE										SE	(.042)		(.030)		(.043)		(.043)		SE	(.039)		(.023)		(.041)		(.036)		(.041)		(.036)		(.041)				
Add. Learning: X ₂	Coef													1.0000		4.9103	.6870			Coef					1.0000		5.9600	.5163			1.0000		5.9600	.5163			
	SE													(.049)		(.109)		(.109)		SE					(.093)		(.207)			(.093)		(.207)		(.093)			

APPENDIX B

VARIABLES IN THE STUDY, VARIABLES RETAINED FOR ANALYSIS AND ESTIMATES FROM
ORIGINAL (NON-INSTRUMENTAL VARIABLES) EQUATIONS

GRADE 6: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

A. Family Background Variables: Z

	1	2	3	4
Born				
20 This City	4.1448			
21 This State	3.8813			
22 Another State	2.7111		- 2.2208	
23 Puerto Rico				
24 Mexico			- 6.2680	
25 Canada	-16.008	-16.4656	-17.0250	- 9.5933
26 Other Country				
27 Do Not Know				
Race				
29 Negro	- 5.7763	- 4.7581	- 5.9865	- 4.1190
30 White		2.5329	1.6160	3.0032
31 American Indian				
32 Oriental				5.1040
33 Puerto Rican	-15.0028	-13.2767	-14.7593	-10.0405
34 Mexican	- 9.1334	- 6.9223	- 6.2209	- 3.9285
35 Other				
Family Size				
37 2				
38 3	3.3991	3.2062	7.8623	6.2156
39 4			4.2698	3.0731
40 5	4.3430	4.2077	8.8943	5.5694
41 6	2.7878	3.6874	5.3372	3.9635
42 7			4.9541	
43 8				
44 9				- 3.1572
45 10				
46 ≥ 11				
Children in Home				
48 1			- 5.8664	
49 2	8.8347	4.2883	3.3220	5.2447
50 3	4.9735			2.8705
51 4	6.2562			4.7925
52 5	6.9057	5.2255	3.7565	7.4277
53 6	9.8637	5.9812	8.5286	9.8944
54 7	7.3856	4.5133	7.0309	9.9679
55 8	8.4250			7.3494
56 9				4.2247
57 ≥ 10		- 5.0552		

A. Family Background Variables: 2 (Cont.)

	1	2	3	4
Acts as Father				
59 Father, Home		2.3714	1.8615	2.3935
60 Father, Away				2.4321
61 Step		2.9293		
62 Foster				
63 Grand	- 7.9644			
64 Relative				
65 Other Adult	- 4.9347			
66 No One				
Acts as Mother				
68 Mother, Home	6.7912	7.8515	8.7327	4.7402
69 Mother, Away				
70 Step				
71 Foster				
72 Grand				
73 Relative	7.7614	15.3001	11.7505	10.1846
74 Other Adult				
75 No One	- 9.6926	- 7.9060	-10.6919	- 6.7409
Father's Education				
77 Some Grade School	- 4.8416	- 2.6447	- 7.9739	- 2.7149
78 Complete Grade School			- 4.4291	
79 Some High School	- 2.5774			
80 Graduated from High School		2.7235		3.5627
81 Voc. or Business School	- 4.8518			
82 Some College		3.5705	4.5505	4.3550
83 Graduated College		6.9121	9.0967	6.8978
84 Graduate School		7.3000	6.7225	7.0001
85 Do Not Know				
Father's Occupation				
87 Technician	- 8.9309	- 5.9714	-10.0262	
88 Official		2.7439	5.3661	5.1033
89 Manager		6.6605	8.5278	6.2432
90 Clerk				2.5868
91 Salesman	- 3.5154			5.3938
92 Farm Owner		3.1106	3.8878	6.3737
93 Farm Worker	- 1.5716			3.5118
94 Factory Worker				2.1803
95 Professional	4.4982	7.9311	10.8317	11.3196
96 Skilled	2.0112	3.8961	3.1658	4.9823
97 Do Not Know	- 2.2275	- 2.1320	- 2.6936	

A. Family Background Variables: Z (Cont.)

	1	2	3	4
Mother Born				
99 This State	3.8827	4.5546	7.5024	7.1548
100 Another State	7.3625	7.0382	10.7146	8.4497
101 Puerto Rico	10.5944	4.3581	7.3378	6.8954
102 Mexico	7.1953			
103 Canada				
104 Other Country				3.8665
105 Do Not Know				
Mother's Education				
107 Some Grade School	- 3.2717	- 3.2757		
108 Complete Grade School	- 5.2620	- 5.5814	- 2.6202	- 6.1736
109 Some High School		- 3.0861		
110 Complete High School	2.8005	2.5929	4.7452	
111 Voc. or Business School				
112 Some College		3.5697		
113 Graduate College		3.0856		
114 Graduate School	- 6.4515	- 5.9415	- 8.8056	- 9.3699
115 Do Not Know				
Mother Employed				
117 Full Time	- 2.6040	- 2.0919		- 1.5396
118 Part Time			- 2.6759	- 1.9859
119 No				
Foreign Language in Home				
121 Yes	3.1260			1.4198
122 No				
Foreign Language Student				
124 Yes				- 2.4322
125 No				
Schools Attended				
176 1		- 2.3103		- 1.5762
177 2	2.2630	- 2.8758		- 2.3645
178 3				
179 4	3.7810		5.1135	
180 ≥ 5		- 6.7032	- 3.2359	- 6.8493
Mothers Interest				
188 Best			- 2.3320	
189 Above Middle	4.5581	6.0925	3.0149	4.2091
190 Middle				
191 Pass				- 4.6111
192 Do Not Know	3.3336			

A. Family Background Variables: Z (Cont.)

	1	2	3	4
Fathers Interest				
194 Best				1.7778
195 Above Middle	- 2.3819			2.9076
196 Middle	- 5.9174		-11.4503	
197 Pass		- 4.0193	- 6.5037	
198 Do Not Know				
R ² =	.7923	.8517	.8184	.8222

GRADE 6: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

B. School Characteristics: X_1

		1	2	3	4
White Students					
182	None	-9.8059		-12.4990	-10.2313
183	Few	-4.1762	2.4937	-9.7858	-9.7751
184	Half	-6.9974		-12.6784	-10.4083
185	Most	-.9722	10.1004		-.7154
186	All		11.1617		
Non White Teacher					
200	Yes	14.6336	-4.8089	-5.0971	7.5756
201	No	16.8306			10.8080
Kindergarten					
209	Yes	-21.6135	4.6541		3.4434
210	No	-25.2164		-4.3340	
Nursery School					
212	Yes		10.4918	5.0822	9.6722
213	No	2.4951	5.2974		5.1621
214	Not Sure			-3.6837	
Time To School					
216	>10 Minutes	9.8158	-8.6058	10.1535	-19.9827
217	20 Minutes	9.1584	-7.8281	10.7679	-18.3496
218	30 Minutes	7.9332	-8.7808	8.0527	-19.0695
219	45 Minutes		-16.3520		-25.3663
220	>60 Minutes		-15.9805		
Transportation					
222	Automobile	9.6475	7.2239	11.5765	12.1027
223	Walk	11.5365	8.5181	13.2277	12.4428
224	School Bus	12.7883	9.6912	15.6116	13.4744
225	Other Bus				
226	Other		-5.9149		
Closer School					
228	Yes	6.6696			8.0227
229	No	7.2192	2.3575		10.4502
230	Do Not Know		-7.9149	-12.2430	

GRADE 6: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

C. Additional Learning Variables: X_2

		1	2	3	4
Pre School Reading					
127	No	- 5.3196	- 7.2512	-62.7043	- 6.5279
128	Occasionally	- 2.5753		-52.6070	
129	Frequently		5.7121	-44.2563	5.5701
130	Often	- 5.6833		-54.2577	
131	Do Not Know			-50.1261	
T.V.					
133	Yes				
134	No			2.2063	2.1080
Telephone					
136	Yes	- 1.4185			
137	No				
Record Player					
139	Yes	3.4984	-11.8610	-21.3095	-18.9101
140	No		-14.4883	-27.3319	-21.8979
Refrigerator					
142	Yes		3.0648	3.4732	2.7673
143	No				
Dictionary					
145	Yes		-24.9667	15.5980	-18.6264
146	No	- 2.2131	-27.4201	12.6398	-21.1899
147	Do Not Know	-11.5864	-37.1193		-29.4343
Encyclopedia					
149	Yes	9.3872	10.3309	12.7567	8.8232
150	No	7.8603	9.8802	10.5934	8.0395
151	Do Not Know				
Automobile					
153	Yes	3.3593	3.1131	4.3972	3.3253
154	No				
Vacuum Cleaner					
156	Yes	5.3409	34.8547	53.4584	26.3965
157	No		28.0703	50.6957	21.8752
Newspaper					
159	Yes	2.3259	1.1817		
160	No				

C. Additional Learning Variables: X_2 (Cont.)

	1	2	3	4
Books Read in Summer				
162 No	- 2.2081	- 5.0161	- 7.2651	- 2.0379
163 1 or 2	- 1.6664	- 2.6555	- 4.1493	
164 About 5				2.6910
165 About 10				
166 ≥ 10				
Hours Watch T.V.				
168 None		3.6755		2.9146
169 $\frac{1}{2}$ Hour	- 6.3258		- 3.6401	
170 1 Hour		3.7839		8.4397
171 $1\frac{1}{2}$ Hours		4.1613	5.5100	6.1689
172 2 Hours		6.9278	5.5366	6.1521
173 3 Hours		3.7031	5.4390	5.6272
174 ≥ 4 Hours				
White Friends				
203 None	- 6.9315	- 7.1658	- 8.8408	
204 Few	- 1.6272	- 5.1046	- 6.0485	
205 Half	- 5.2986	- 9.4352	-11.2282	
206 Most		1.4470	1.6426	8.4857
207 All				6.9227
Talk About School				
256 Every Day		- 2.3274	- 1.5750	
257 Twice a Week	2.0813			1.9242
258 Occasionally	6.8535	5.2855	7.1210	5.2006
259 Never				
$R^2 =$	.7399	.8133	.7846	.7864

GRADE 9: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

A. Family Background Variables: Z

	1	2	3	4	5
Area Most of Life					
24 This City		-2.6019			
25 This State					
26 Another State			3.1496		
-27 U.S. Possession	-33.6937		-26.2874	-29.1102	-43.4470
28 Mexico					
-29 Canada	-46.6385			-70.1295	-31.2474
30 Another Country					
Community					
32 Country					
-33 Small Town	2.3181	1.7865	1.8996		
-34 Medium Size City	2.6618		1.8839		
35 Suburb Medium Size City				-6.0567	
-36 Large City	4.1982				
-37 Suburb Large City	5.7615		5.8760		
-38 Very Large City	5.0970		4.2699		
39 Suburb Very Large City					
Race					
-41 Negro	-5.3752	-8.2710	-4.9415	-6.9277	-6.2089
42 White		2.7567	4.9993	3.3014	
43 American Indian					
44 Oriental					
45 Puerto Rican		-8.3950			
-46 Mexican	-6.5155	-10.4957	-5.6665	-7.3556	-8.5424
47 Other					
Number in Home					
49 2					
50 3		4.7789			
51 4					
52 5			4.6792		
53 6					
54 7					
55 8					
56 9					
57 10					-11.3457
58 >11				-13.1153	-5.6852

A. Family Background Variables: Z (Cont.)

		1	2	3	4	5
	Child in Family					
60	None		-9.6806	-8.3102	-15.6254	-7.7337
61	1					
62	2					
63	3					
64	4					-5.7367
65	5					-6.2629
66	6		5.6761		7.3517	
67	7			-5.1766	-6.5030	
-68	8	-4.1350				
69	>9					
	Older Siblings					
-71	None	6.2627	8.5456	8.2110	14.0914	
-72	1	6.1933	3.8459		6.4441	
73	2				6.1246	
-74	3	4.7755				
75	4					
-76	5	7.2825				
77	6		-10.7370	-7.2790		-9.4012
78	7					
79	8					
80	>9					
	Number of Dropouts					
82	Does Not Apply					5.6741
83	None				-5.3455	
84	1					
85	2					
86	3					
87	4			-10.5507		
-88	5	-10.3219			-15.1119	
89	6					
-90	7	-21.7567				
91	>8					
	Foreign Language Parent					
93	Yes					
94	No		4.6850			
	Foreign Language Student					
96	Frequently					
-97	Occasionally	10.3791	7.7142			6.3112
98	Rarely		6.5369	6.7494	5.0376	4.7641
99	No					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Rooms in Home					
101 1		-19.0302			
102 2					
103 3					
104 4					
-105 5	-3.6130	-4.3081	-3.4445	-8.7790	-4.7683
106 6					
107 7					
108 8					
109 9					
-110 >10	9.0432	7.9658	8.9702		
Acts as Father					
112 Father Home					
113 Father Away					
114 Step					
115 Foster					
-116 Grand	-12.2470	-15.4312		-12.1605	-15.1102
117 Relative		-11.9628		-19.2514	
118 Other Adult				26.7104	
119 No one					
Acts as Mother					
-121 Mother Home	5.6237	4.0086	6.2128	15.7538	8.7387
122 Mother Away					
123 Step					10.7551
124 Foster					
-125 Grand	12.5620			15.8221	20.9036
126 Relative				28.1043	20.7114
127 Other Adult					
128 No One		-19.3102	-12.3099		
Fathers Occupation					
130 Technical					
131 Official					
132 Manager			4.6549		
133 Semi-skilled		4.8492	4.5135	7.0945	5.6413
134 Salesman					
-135 Farm Owner	2.9603	5.3503	4.1193	5.6492	7.7167
136 Farm Worker					
-137 Laborer	3.9573	4.6956	4.4349	6.3604	5.2055
138 Professional			6.3155	9.1404	
-139 Skilled	2.8009	5.8806	3.4537	6.2137	4.4126
140 Do Not Know					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Fathers Education					
142 Some Grade School					
-143 Complete Grade School	4.5680	5.9364	6.9176	6.8843	7.8230
144 Some High School		-2.6504			
145 Graduated, High School			4.2838		4.6161
146 Voc. or Bus. School					
147 Some College					
148 Graduated, College					11.3320
149 Graduate School					18.2241
150 Do Not Know				-3.7897	
Mothers Education					
152 Some Grade School					
153 Complete Grade School					
154 Some High School					
-155 Graduated, High School	3.1989	5.5235	4.5908	7.2719	5.1505
-156 Voc. or Bus. School	10.6149	20.1190	20.6949	17.2560	16.4714
157 Some College			6.9241	7.8873	6.5455
158 Graduated, College		12.7611	10.6383	9.6591	
159 Graduate School					
160 Do Not Know					
Mothers Community					
162 Country	1.9874	-1.7561			5.1920
163 Small Town					6.5424
164 Medium Size City					3.9878
165 Suburb Medium Size City					
166 Large City					
167 Suburb Large City					
168 Very Large City					7.8679
169 Suburb Very Large City					
170 Do Not Know		-6.2918	-4.5047		
Source of Income					
-172 Father	3.8125		3.3023		
173 Mother					
174 Stepfather		-13.0687		-16.9098	-22.4776
-175 Stepmother	-14.6580				-22.0354
-176 Other	6.4581		7.8466	7.0968	
177 Do Not Know					-13.4821
Mother Employed					
179 Full Time		-2.2743			
180 Part Time					
181 No					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Mothers Interest					
-183 Best	-6.2864		-5.9457	-10.3088	-6.1536
184 Above Average		9.4479			
185 Average			-9.5144		
186 Passing					
187 Do Not Know					
Fathers Interest					
189 Best					
190 Above Average					
191 Average					
192 Passing					-8.9544
193 Do Not Know					
Fathers Desires					
-200 Does Not Care	-16.0574	-22.4569	-17.4709	-18.1713	
-201 High School	-10.2916	-9.4612	-12.4065	-12.1067	-10.8943
202 Technical		-5.6791	-4.9283		-3.8233
-203 Some College	-4.8265	-5.5890	-10.6460		
204 Graduated, College					
205 Graduate					
206 Not Home					
207 Do Not Know					
Mothers Desires					
-209 Does Not Care	-20.6339		-16.3983	-27.5840	-30.2259
210 High School					
211 Technical					
212 Some College				-11.8555	-7.3685
213 Graduated College		3.0198		4.3625	
214 Graduate		9.0186	8.3260	15.0277	
215 Not Home		25.8799	20.1812	19.4287	
216 Do Not Know					
P.T.A. Attendance					
-218 None	5.3913	6.3505	6.6649	8.4670	7.7236
-219 Occasionally	3.5050	5.9453	6.9515	6.1154	4.5678
220 Half					
-221 Most	4.6155	5.7550	5.3355	7.9882	6.2840
-222 Does Not Exist	3.3105	5.5833	6.6561	6.7867	6.8223
223 Do Not Know					
Change School					
291 Never					
-292 1	-3.9197			-3.2367	
293 2					
294 3					
295 >4					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
	Last Change Schools				
-297 Never	-2.1532				
298 <Year					
299 Year		-7.5221			
300 2 Years					
301 3 Years			-4.8789		
302 4 Years					
303 >5 Years					
R ² =	.8864	.9398	.9444	.9111	.8969

GRADE 9: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

B. School Characteristics: X_1

	1	2	3	4	5
School Curriculum					
275 General		2.8766			
276 College Prep					
277 Commercial	-4.4094	-7.9153	-6.7694	-5.0426	-7.0733
278 Vocational	4.4934				
279 Agriculture				6.9455	
280 Arts					
281 Other					
282 Do Not Know	-1.1904				
Kindergarten					
284 Yes	2.7700	2.4005	2.8574	1.8970	2.8988
285 No					
Nursery School					
287 Yes	5.2106	14.4152	14.9597	17.1071	13.9415
288 No	5.7559	11.1057	7.7290	11.5385	13.5983
289 Do Not Know					
White Classmates					
205 None		-6.7094	-5.9982	-4.9763	-4.4037
306 <Half	3.0258				
307 Half	-2.7068	-7.7255	-6.4196	-6.4623	-4.2400
308 Most					
309 All					
White Teacher Class					
311 None					
312 <Half			-5.0982		
313 Half					
314 Most					
315 All					
White Student School					
317 None					
318 <Half					
319 Half					
320 Most	2.5264	4.4448			
321 All					
White Teacher School					
323 None	-8.0988	-6.7955	-6.7400	-8.6238	-4.9002
324 <Half					
325 Half	-7.3512	-15.9855	-12.6787	-18.1406	-17.3451
326 Most		-2.1878			
327 All					

B. School Characteristics: X_1 (Cont.)

		1	2	3	4	5
	Future Teacher Club					
349	Yes	-10.0932	-11.4501	-13.0254	-19.8221	-11.5954
350	No					-1.5323
351	Does Not Exist					
	Different Ract					
353	>Third					
354	4-6		-5.0876	-4.5667		-4.0827
355	7-9					
356	None	2.3812	2.5850			
	Athletic Team					
358	Yes	3.3325	6.7479	1.6796		
359	No	2.7315	6.1138			
360	Did Not Exist					
	Student Council					
362	Yes					
363	No	2.9252		3.3635	3.5459	2.6348
364	Did Not Exist					
	Debating Club					
366	No					-4.3431
367	Active		2.1704	4.4601	2.9609	
368	Belonged	5.3041	6.0159		7.1945	
369	Did Not Exist					
	Hobby Club					
371	No	-2.0505	-3.0911		-4.9648	
372	Active					
373	Belonged		-7.8925		-9.8407	
374	Did Not Exist			3.8725		
	Science Courses					
376	Yes	.9295				
377	No					
	Foreign Language Course					
379	Yes			3.1563	2.4699	
380	No		-1.7629			
	Social Studies Courses					
382	Yes	-.8545		-1.4287	-1.1722	-1.9789
383	No					
	English Course					
385	Yes	2.6858	6.3729	5.7864	5.6893	6.7400
386	No					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Math Course					
388 Yes	2.8947	3.4791	3.3611	3.4779	4.4538
389 No					
Industrial Arts Course					
391 Yes		2.8774		2.6662	
392 No					
Vocational Course					
394 Yes		-4.5503	-3.9453		-3.1618
395 No					
Commercial Course					
397 Yes					
398 No					
Agriculture Course					
400 Yes				-3.6807	
401 No					
Home Economics Course					
403 Yes	-2.4735	-4.8446	-2.9054	-2.3972	-3.6968
404 No					
English Grade					
406 A				3.4351	
407 B					
408 C		-3.3939			
409 D					
410 Failed				16.8346	
411 Did Not Have		-16.6419	-15.4445	18.5970	-19.8235
Math Grade					
413 A					
414 B	-2.5873	-4.3693			
415 C		-4.9117			
416 D		-4.7997			
417 Failed	-6.5540	-13.9451	-8.9877	-11.9319	-15.6292
418 Did Not Have					
English Ability					
420 Highest			3.3470		
421 Middle				-4.9775	
422 Lower				-7.7066	
423 Does Not Have					
424 Do Not Know					
$R^2 =$	.8823	.9072	.9176	.8898	.8767

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Repeat English					
426 Yes	-3.8787	-7.2831	-6.7767	-9.6807	-5.3482
427 No					
Average Last Year					
429 A			7.5691		
430 B		5.3760	5.7040	3.4082	
431 C	3.4460	11.5340	11.7064	6.7819	3.7484
432 D	-6.5729				-7.9310
433 Do Not Know	-3.1996				
See Counselor					
442 Yes	2.0210	2.8611	3.4360	3.3596	7.5749
443 No	-2.2980		-3.2748		
444 Does Not Exist				-6.5062	4.2595
Vocational Program					
446 Enrolled	-5.2694	-4.1830	-7.7903	-10.3322	-4.7601
447 Like To		-3.1874	-5.4443		
448 No					3.1791
Time To School					
454 10 Minutes	2.2784	2.7409	8.8607	9.7715	
455 20 Minutes	3.7579		7.4802	7.7038	-2.5998
456 30 Minutes			6.2466	9.4172	
457 45 Minutes	4.7340		8.7612	9.5765	
458 >60 Minutes	-2.9023	-4.3933			-10.4499
Transportation					
460 Automobile					
461 Walk			-1.9979	-2.5379	-2.2423
462 School Bus		2.0879			
463 Other Bus	2.1794				
464 Other					
8th Grade School					
490 Public			-3.9653		
491 Catholic					
492 Jewish	-7.2045	-14.6814	-15.0401	-13.1877	-13.2456
493 Protestant			-13.1372	-12.3442	
494 Military					

GRADE 9: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

C. Additional Learning Variables: X_2

	1	2	3	4	5
Pre School Reading					
225 No		- 8.7049	- 5.5962	-14.8412	
226 Occasionally		7.7550	7.0991		9.3186
227 Frequently	4.6925	13.8186	11.8595	10.6483	13.8129
228 Often	- 8.1082			-10.1247	
229 Do Not Know					
Television					
231 Yes					
232 No					
Telephone					
234 Yes			2.5801		
235 No					
Record Player					
237 Yes	4.0589				
238 No					
Refrigerator					
240 Yes		6.7857	5.6055	12.4828	8.3127
241 No					
Dictionary					
243 Yes					
244 No			- 4.9335		
Encyclopedia					
246 Yes				4.0657	
247 No	- 1.8147				
Automobile					
249 Yes	2.3469	9.6422	7.5262	8.4236	7.5700
250 No					
Vacuum Cleaner					
252 Yes	6.1859	9.5028	10.6483	9.4027	8.7614
253 No					
Newspaper					
254 Yes					
255 No					
Public Library					
257 $\geq$ Weekly					
258 $\geq$ Monthly				- 5.3868	- 3.8445
259 $\leq$ Monthly	2.5073	3.7887	6.1239	8.0102	4.9929
260 Never	2.1473	2.7042			

C. Additional Learning Variables: X_2 (Cont.)

	1	2	3	4	5
Number of Magazines					
263 None		8.1968	10.2565	8.8672	7.1225
264 ≤ 2	5.3018	10.0494	11.1057	9.8658	10.2300
265 ≤ 4	5.8960	14.0386	15.0634	13.6719	13.2690
266 ≤ 6	5.3052	18.5062	16.4865	15.0643	16.0497
267 ≥ 7					
Number of Books					
269 0 to 9		-11.6853	- 9.0153		-17.8044
270 10-24	-10.5252	-21.8103	-24.9479	-19.0731	-27.1303
271 25-99	- 3.6472	-13.4178	-15.4049	- 9.3636	-20.6879
272 100-249		- 7.4164	-11.5106		-17.3576
273 ≥ 250					-12.3900
Books Read in Summer					
334 None				- 5.0666	
335 1 to 5		- 2.9662		- 4.8955	
336 6-10					
337 11-15		- 7.0690		- 9.7146	
338 16-20					
339 ≥ 20					
Hrs. Watch T.V.					
341 None		9.6452	7.6503	11.3772	10.6503
342 $\frac{1}{2}$ Hr. Per Day	- 8.1238			- 8.8935	
343 1					
344 $1\frac{1}{2}$	7.5532	19.1053	14.9219	22.2300	15.2417
345 2	3.7395	17.5284	14.4078	19.1968	14.5196
346 3	4.3022	15.4077	9.5667	16.2005	11.8850
347 ≥ 4					
School Discussions					
195 Daily		-16.8540	-14.0013	-18.7317	-10.7129
196 Weekly		-22.4958	-20.6395	-25.3734	-16.0307
197 Monthly		-14.5370	- 8.8067	-16.9085	- 8.0105
198 Hardly	13.1389				
Outside Job (included in 74)					
435 None	2.7618	4.5679	6.8403	6.7160	5.0273
436 1-5 Hrs.					
437 6-10 Hrs.					
438 11-15 Hrs.					
439 16-20 Hrs.					
440 ≥ 21 Hrs.		8.6005	12.9346		8.2306
$R^2 =$	.8245	.8814	.8905	.8492	.8262

GRADE 12: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

A. Family Background Variables: Z

	1	2	3	4	5
Area Most of Life					
24 This City					
25 This State					
26 Another State					
27 U.S. Possession				-53.0965	-62.1088
28 Mexico					
29 Canada					
30 Another Country	-14.4157				
Type of Community					
32 Country			-2.4443		
33 Small Town					
34 Medium Size City					
35 Suburb Medium Size City					
36 Large City					
37 Suburb Large City					
38 Very Large City				6.6378	
39 Suburb Very Large City					
Race					
41 Negro	-5.5421	-6.4285	-5.6264	-6.8999	-12.6728
42 White		3.9931	5.4130	6.9773	
43 American Indian					
44 Oriental					
45 Puerto Rican	-7.5218	-15.8503	-7.4902		
46 Mexican	-7.0787	-9.9679			
47 Other					
Number in Home					
49 2					15.5600
50 3		8.9588	5.0096	5.4164	8.7325
51 4	5.2401	7.0533	4.9975	7.4159	9.1213
52 5			4.2901	8.8259	7.6859
53 6					
54 7					
55 8					8.0257
56 9					
57 10					
58 >11					
Children in Family					
60 None		-13.1318	-8.5126	-11.1330	-11.8906
61 1	-4.1014				
62 2					
63 3			4.4219		
64 4		5.0451			
65 5					
66 6					

A. Family Background Variables: Z (Cont.)

		1	2	3	4	5
	Children in Family					
67	7					
68	8			-8.8390		
69	9 or More					
Older Siblings						
71	None	10.4270	8.3169	8.5624	9.6744	
72	1	10.5889	5.5822	5.1023	10.6604	5.5686
73	2	5.6060				
74	3	8.4662		7.8999	13.7348	
75	4	4.1200				-9.3280
76	5					
77	6					
78	7					
79	8					
80	9 or More					
Number of Dropouts						
82	Does Not Apply					4.3986
83	None				-4.3058	
84	1					
85	2					10.0756
86	3					
87	4					
88	5					
89	6	-21.4122	-24.4693		-29.9381	
90	7					
91	8 or More					
Foreign Language Parents						
93	Yes	3.8175				
94	No					
Foreign Language Student						
96	Frequently					-13.4801
97	Occasionally			-5.6203		
98	Rarely		3.4627	7.6699	5.7875	4.5748
99	No					
Rooms in Home						
101	1					
102	2					
103	3					
104	4	-3.5514				-5.7898
105	5					
106	6	-2.4237	-5.2566			-5.3663
107	7					
108	8		6.6368			
109	9					
110	10 or More			10.0538		

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Acts as Father					
112 Father, home					9.5087
113 Father, away	-4.6673		-7.6514		
114 Step	4.7847				
115 Foster					
116 Grand	-9.6019	-15.2322			
117 Relative					
118 Another Adult		-31.5866	-24.5340		
119 No One					
Acts as Mother					
121 Mother, home					
122 Mother, away	-7.7610				
123 Step					-15.0144
124 Foster					
125 Grand				-16.3514	12.2364
126 Relative					
127 Another Adult					
128 No One	8.3324				
Fathers Occupation					
130 Technical					
131 Official	8.7284				
132 Manager	5.1848			6.4158	
133 Semi-skilled	8.9254	10.4249	5.3074	14.7930	4.9495
134 Salesman					
135 Farm Owner	5.4470	4.5539	4.9494	11.1170	
136 Farm Worker			-5.8793		
137 Laborer	6.5782	4.1673	7.7966	8.4681	
138 Professional	5.2710				
139 Skilled	8.1124	8.5534	6.6441	13.0998	
140 Do Not Know			-6.9869		
Fathers Education					
142 Some Grade School					
143 Completed Grade School		4.0274			
144 Some High School					
145 Graduated High School					
146 Tech. or Bus. School					
147 Some College					
148 Graduated, College	8.3866	16.0803	19.2199	22.9038	
149 Graduate School		9.4838			13.4269
150 Do Not Know					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Mothers Education					
152 Some Grade School					
153 Completed Grade School					
154 Some High School			-3.6051		
155 Graduated High School	3.4123	5.1131	4.8344	7.7996	7.1194
156 Tech. or Bus. School		11.7035	9.8659	14.1476	15.9964
157 Some College	4.7615	7.0379	7.0948	11.1928	11.9426
158 Graduated College					
159 Graduate School					
160 Do Not Know					
Mothers Community					
162 Country					
163 Small Town					
164 Medium Size City					
165 Suburb Medium Size City	-7.6750				
166 Large City					
167 Suburb Large City		14.2496			
168 Very Large City			9.5517		
169 Suburb Very Large City					
170 Do Not Know			-5.6867		-7.3042
Source of Income					
172 Father					
173 Mother					
174 Step Father					
175 Step Mother		-40.7162	-40.6355	-42.3865	-26.5025
176 Another					
177 Do Not Know					
Mothers Employed					
179 Full Time	-2.3117	-3.6238			-4.2461
180 Part Time					
181 No Job					
Mothers Interest					
183 Best					
184 Above Average			5.9116		7.8600
185 Average					
186 Passing	-9.9500				-31.7295
187 Do Not Know					
Fathers Interest					
189 Best	-4.1738	-6.0598	-9.1908	-10.0320	-8.8121
190 Above Average			-8.2927		-11.6639
191 Average	3.7948				
192 Passing			-16.5288	-23.1927	
193 Do Not Know					

A. Family Background Variables: Z (Cont.)

	1	2	3	4	5
Fathers Desires					
200 Doesn't Care				-14.8164	
201 High School			-6.6769	-11.7454	
202 Tech. or Bus. School					
203 Some College			-5.5008	-11.1556	
203 Graduate College	1.8884	6.2867	6.4105	4.2319	6.9480
205 Graduate School	4.8662				
206 Not Home					
207 Do Not Know					
Mothers Desires					
209 Doesn't Care	-13.4562				
210 High School	-3.1114	-6.6934			
211 Tech. or Bus. School					
212 Some College					
213 Graduate College					
214 Graduate School		5.0901	11.7627	7.3521	7.3671
215 Not Home				32.1205	
216 Do Not Know					
P.T.A. Attendance					
218 None			5.5801		-2.8964
219 Occasionally			8.0956	4.2882	
220 Half		-5.6653			
221 Most			5.1334		
222 Does Not Exist			7.1192	2.2904	
223 Do Not Know					-5.9797
Change Schools					
294 Never					
295 1	-4.3161	-3.0671	-3.2192	-5.4073	
296 2					
297 3					
298 4 Times or More					
Last Changed Schools					
300 Never	-2.0699				
301 Less than Year					
302 1 Year	-8.6546			-11.5102	
303 2 Years					
304 3 Years					
305 4 Years					
306 5 Years or More					
R ² =	.8907	.9133	.9282	.9022	.8940

GRADE 12: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

B. School Characteristics: X_1

	1	2	3	4	5
School Curriculum					
275 General	3.7473	2.1988	4.8847	3.7780	
276 College Prep	3.8804	3.5757	6.7273	5.8120	
277 Commercial					
278 Vocational	3.8966			5.9258	
279 Agriculture					
280 Arts	3.6562				
281 Other					
Choice Curriculum					
283 My Choice			2.4204		
284 Assigned					-5.1296
285 One Program					
Kindergarten					
287 Yes	2.3044	1.6841	2.7399	1.8137	
288 No					-1.5137
Nursery School					
290 Yes	6.6401	5.5293	12.4941	4.9921	
291 No	5.0038		5.1785		
292 Do Not Know					
White Student Class					
308 None		-9.4411			
309 Less Than Half	-2.7282	-5.9523	-7.9972	-7.1336	
310 Half					
311 More Than Half					
312 All		3.1694			7.9368
White Teacher Class					
314 None					
315 Less Than Half			1.4898		
316 Half	-1.6729			-4.4537	
317 More Than Half				-3.2219	
318 All		.8373	2.0371		
White Student School					
320 None	-3.5758	-11.9177	-14.2563	-13.6149	-14.5188
321 Less Than Half		-5.5829	-8.2343	-5.2299	-5.4192
322 Half		-7.0611	-4.9883		
323 More Than Half					
324 All					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
White Teacher School					
326 None	-3.2472				
327 Less Than Half			4.9630	4.5439	
328 Half	-5.7499	-11.7941	-8.6209		-18.7003
329 More Than Half	-1.1865		-1.9800	-3.2592	-4.1249
330 All					
College Catalog					
332 Yes	1.9157	3.0092	3.4133		4.9064
333 No				-4.6331	
Different Race					
356 1-3 Grade	-4.2475				
357 4-6 Grade		6.0085		8.7759	6.2313
358 7-9 Grade		8.5268		7.2223	5.2529
359 10-12 Grade					
360 Never					
Teacher Club					
362 Yes	-1.8882	2.5965	2.1112		9.1685
363 No	-1.9532	-2.0270		-2.4922	4.4467
364 Does Not Exist					6.8063
Athletic Team					
366 Yes					
367 No					
368 Does Not Exist					
Student Council					
370 Yes					
371 No	1.5393				
372 Does Not Exist					
Debating Clubs					
374 No			-1.2822		-4.5835
375 Active					-6.4856
376 Belonged		-1.7446			-5.1707
377 Does Not Exist					-4.0200
Hobby Clubs					
379 No					-1.9111
380 Active					
381 Belonged		2.7535			
382 Does Not Exist					-2.1471

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Science Courses					
384 None	3.1079				
385 $\frac{1}{2}$ Year					
386 1 Year			-2.7249		
387 $1\frac{1}{2}$ Years					
388 2 Years				8.4089	
389 $2\frac{1}{2}$ Years					
390 3 Years			19.6810		
391 $3\frac{1}{2}$ Years	3.9878				
392 4 Years					
393 More Than 4 Years					
Foreign Language Courses					
395 None					
396 $\frac{1}{2}$ Year					
397 1 Year					
398 $1\frac{1}{2}$ Years					
399 2 Years					
400 $2\frac{1}{2}$ Years		-13.1033			-4.6294
401 3 Years		-16.0913			-7.4937
402 $3\frac{1}{2}$ Years					
403 4 Years		-12.2362			
404 More Than 4 Years					
Social Studies Course					
406 None				-15.1729	
407 $\frac{1}{2}$ Year	3.0756			-5.4950	
408 1 Year					
409 $1\frac{1}{2}$ Years	5.8383	4.5589	3.5150		
410 2 Years					
411 $2\frac{1}{2}$ Years	5.9025	4.6449	6.0115		2.9400
412 3 Years					
413 $3\frac{1}{2}$ Years	6.0857	5.3381	4.6943		
414 4 Years	5.2453	4.1402	3.5987		
415 More Than 4 Years					
English Courses					
417 None	-9.6261	-20.5288	-20.5771	-31.8400	
418 $\frac{1}{2}$ Year					
419 1 Year					
420 $1\frac{1}{2}$ Years	-11.9147				
421 2 Years				-10.3597	
422 $2\frac{1}{2}$ Years	-11.1637				
423 3 Years					
424 $3\frac{1}{2}$ Years					
425 4 Years					
426 More Than 4 Years					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Math Courses					
428 None					
429 $\frac{1}{2}$ Year	-9.7654	-9.0345			-9.4879
430 1 Year					
431 $1\frac{1}{2}$ Years					
432 2 Years					
433 $2\frac{1}{2}$ Years					
434 3 Years		-3.0739			
435 $3\frac{1}{2}$ Years					
436 4 Years	2.4412		2.3284		5.6426
437 More Than 4 Years					19.3830
Arts Courses					
439 None	-1.6779	-4.8217	-1.6976		
440 $\frac{1}{2}$ Year					
441 1 Year					
442 $1\frac{1}{2}$ Years					
443 2 Years					
444 $2\frac{1}{2}$ Years					
445 3 Years					-4.1484
446 $3\frac{1}{2}$ Years					
447 4 Years				5.4953	
448 More Than 4 Years					
Vocational Courses					
450 None					
451 $\frac{1}{2}$ Year					
452 1 Year					
453 $1\frac{1}{2}$ Years					
454 2 Years					
455 $2\frac{1}{2}$ Years					
456 3 Years					
457 $3\frac{1}{2}$ Years					
458 4 Years					
459 More Than 4 Years			28.2160	31.7799	
Commercial Courses					
461 None	-1.7066		-2.0237	-2.1152	
462 $\frac{1}{2}$ Year					
463 1 Year	-1.6775				
464 $1\frac{1}{2}$ Years					
465 2 Years					
466 $2\frac{1}{2}$ Years					
467 3 Years					
468 $3\frac{1}{2}$ Years					
469 4 Years				8.4712	
470 More Than 4 Years					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Agriculture Courses					
472 None		2.9386	4.7410	4.5818	
473 $\frac{1}{2}$ Year		-7.6990			
474 1 Year					
475 $1\frac{1}{2}$ Years					
476 2 Years					
477 $2\frac{1}{2}$ Years	-17.1133				
478 3 Years	-1.8991				
479 $3\frac{1}{2}$ Years					
480 4 Years					
481 More Than 4 Years		18.1788		22.0452	
Economic Courses					
None					
483 $\frac{1}{2}$ Year	2.2549	3.1194	3.5166	3.6627	2.8722
484 1 Year					
485 $1\frac{1}{2}$ Years	1.7183				
486 2 Years					
487 $2\frac{1}{2}$ Years		-2.9352			
488 3 Years	10.7570				
489 $3\frac{1}{2}$ Years					
490 4 Years					
491 More Than 4 Years					
English Grade					
494 A					
495 B			-3.1787		
496 C					
497 D					
498 Failed					
499 Did Not Have				-47.1502	-75.2494
Math Grades					
501 A					
502 B				-4.0948	-3.7502
503 C				-4.0789	-6.5315
504 D					
505 Failed					-22.0731
506 Did Not Have					
English Ability					
508 Highest		-3.9063			
509 Middle	-3.8751	-4.6891	-6.0259	-7.9052	-4.1858
510 Lower		6.6158	7.5872		
511 Do Not Have					
512 Do Not Know					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Repeating English					
514 Yes					
515 No		6.3294	7.7315	9.0739	
Average Last Year					
517 A					
518 B					4.1649
419 C					5.5126
420 D					
421 Do Not Know	-2.5071				
See Guidance Counselor					
530 Yes				4.2932	-9.5268
531 No					-6.6540
532 Do Not Have					
Time Guidance Counselor					
534 None					
535 1 Time	-3.7189		-5.1757		-7.2442
536 2-3 Times					
537 4-5 Times	-3.5925	-2.4522	-4.8377		-9.7637
538 6 Or More Times		-8.3518	-10.1132		
539 Does Not Exist					
Encourage Education					
541 Yes, College	1.6060	2.6333	2.1130	3.8849	
542 Yes, Technical					
543 Yes, Business					-2.5771
544 Yes, Other		-7.0947			-9.7155
545 No					
Vocational Course					
547 In Training					
548 Yes	-1.1152				
549 No	-1.5690				

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Type Vocational Program					
551 Agriculture	-15.1626		-34.5326		
552 Air Conditioning				7.9656	32.5003
553 Airplane Mechanics					46.5006
554 Auto Body Mechanics					28.8113
555 Automobile Mechanics					31.1004
556 Brick Masonry				9.6056	26.4680
557 Cabinet Making					36.5968
558 Carpentry					-8.0532
559 Commercial Arts			-7.5713		
560 Cooperative Office Tng.					
561 Cosmetology					
562 Diesel Mechanics					21.6207
563 Distributive Education					
564 Electricity					
565 Food Trades					
566 Foundry					
567 Indust. Coop. Tng.					
568 Machine Shop					
569 Domestic Service					
570 Needle Trade					
571 Painting & Decorating					
572 Plumbing					
573 Practical Nursing					
574 Printing					
575 Radio-TV Repair					
576 Sheet Metal Work					
577 Welding					
578 Other					
Training for Job					
580 $\frac{1}{2}$ Year					
581 1 Year					
582 $1\frac{1}{2}$ Years					
583 2 Years					
584 $2\frac{1}{2}$ Years					
Work Study Program					
586 Yes					
587 No	-0.4447				-0.9403
Job Desired Training					
589 Yes					
590 Not Qualified		-6.6389	-6.4798	-7.6878	
591 Course Full				-9.3052	
592 Did Not Try		-5.9132			
593 Does Not Exist					

B. School Characteristics: X_1 (Cont.)

	1	2	3	4	5
Time To School					
595 10 Minutes Or Less					
596 20 Minutes		-8.6629	-16.0988		-8.5404
597 30 Minutes			-17.3958		
598 45 Minutes			-14.3344		
599 1 Hour Or More		6.5577			
Transportation To School					
601 Automobile				-7.9542	-7.3863
602 Walk					
603 School Bus					
604 Other Bus	-3.4751				
605 Other					
8th Grade School					
607 Public	-3.2694				-5.9011
608 Catholic					
609 Jewish			-9.2075		-9.3711
610 Protestant			-8.4035		
611 Military		5.8562		8.9835	
$R^2 =$	.9130	.9242	.9353	.9292	.9040

GRADE 12: FINAL ESTIMATES FOR RETAINED INDEPENDENT VARIABLES

C. Additional Learning Variables: X_2

	1	2	3	4	5
School Discussion					
195 Daily	- 4.9121	- 6.4020	- 9.5028		
196 Weekly	- 6.0402	- 7.1574	-12.6219		
197 Monthly		- 8.6811	-12.8371		
198 Hardly					
Pre School Reading					
225 No					
226 Occasionally					7.9310
227 Frequently			5.4562	6.5478	10.0224
228 Often	- 7.1460	- 8.1298	- 5.8527	- 6.2302	
229 Do Not Know	- 3.4161	- 6.0448			
T.V.					
231 Yes	- 4.7704	-11.3624	- 8.4596	- 9.0406	-10.1654
232 No					
Telephone					
234 Yes	2.5188	4.2736	4.5015	-25.0648	
235 No				-29.7899	- 2.9811
Record Player					
237 Yes		- 5.4995	- 3.8068		- 9.0137
238 No			12.5190		
Refrigerator					
240 Yes	- 9.7260				-10.9083
241 No					
Dictionary					
243 Yes	6.1413		9.6381	11.6273	
244 No		- 6.2873			
Encyclopedia					
246 Yes					
247 No					
Automobile					
249 Yes	5.2664	8.6045	5.5895	7.4521	14.9155
250 No					
Vacuum Cleaner					
252 Yes	10.0674	16.1207	17.3043	18.8885	16.3397
253 No					
Newspaper					
255 Yes					
256 No					

C. Additional Learning Variables: X_2 (Cont.)

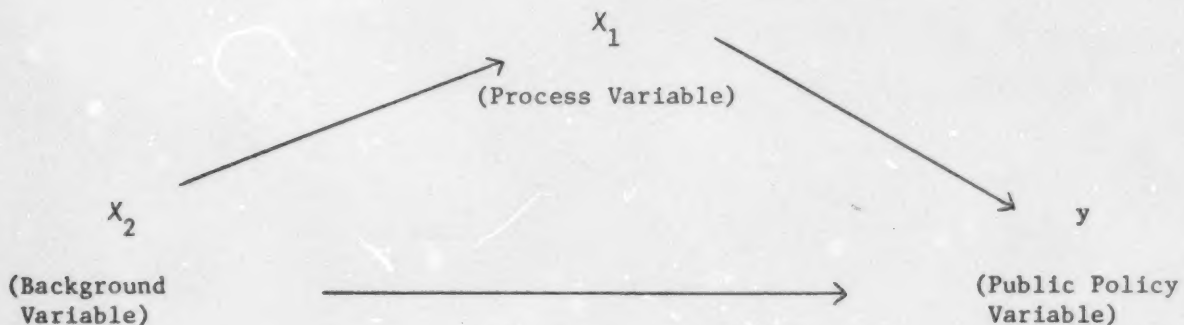
	1	2	3	4	5
Public Library					
258 Once a Week			- 5.9065	- 8.1551	- 4.7016
259 2 or 3 times a Month			3.8278		
260 Once a Month or Less	3.1357				
261 Never					
Number of Magazines					
263 None	- 6.1676				
264 1 or 2		7.0728	6.0885	5.7066	4.8816
265 3 or 4		9.3104	6.0664	7.3691	10.1391
266 5 or 6		11.7737	9.4849	10.7686	15.2287
267 7 or More	- 4.8054				
Number of Books					
269 0-9			- 8.6954		-11.8335
270 10-24	- 3.3527	- 7.6488	-15.8262	- 6.4160	-11.0323
271 25-99	- 2.9524	- 6.6828	-11.4296	- 4.8535	-10.9467
272 100-249			- 8.2497		- 9.2335
273 250 or More					
Summer Reading					
335 None		14.2208	15.6507	16.0528	12.6402
336 1-5 Books	- 3.7653				
337 6-10 Books	-14.7689				-24.2522
338 11-15 Books	3.9412	6.4519		7.4054	
339 16-20 Books			- 7.5417	-12.4936	- 9.6198
340 21 or More Books					
Hrs. Watch T.V.					
342 None		-12.4880	-11.5186	-20.7751	
343 $\frac{1}{2}$ Hour Per Day					
344 1 Hour Per Day					
345 $1\frac{1}{2}$ Hours Per Day					
346 2 Hours Per Day		- 9.5511	- 8.3689		
347 3 Hours Per Day					
348 ≥ 4 Hours Per Day					
White Friends (used only in 6th, not 9th grade)					
350 None					
351 Less than half				-11.2158	
352 Half					
353 More than half					
354 All	2.6360				
Outside Job					
523 None					-10.4962
524 1-5 Hours	- 3.8218	- 6.1663	- 9.5210	-11.2559	-17.3043
525 6-10 Hours		- 5.0959	- 9.7551	-10.8645	-14.0339
526 11-15 Hours					
527 16-20 Hours	5.3813		9.8011		
528 21 or More Hours				- 8.1551	-12.2462
$R^2 =$	.8513	.8554	.8752	.8469	.8277

Appendix C.

Background Variables and Process Variables

Background Variables and Process Variables

In political research we frequently encounter situations in which we want to compare the relative effects of two different classes of independent variables on some dependent variable. One of these classes of variables can be termed "background variables" because they usually refer to measures of factors which we conceptualize as initial influences in a causal sequence. Examples are the socio-economic indicators measures at the individual or aggregate level which we take as indicators of the social background of the individual or the community, state or nation. The other class of variables frequently are measures of some "process" which "intervenes" between the background and dependent variables. For example, political process and other political variables usually are conceptualized as less "initial" in their causal impact on public policy variables and are ordinarily thought to be in turn influenced by the background variables themselves. Thus we frequently hypothesize the following causal sequence in political research:



In research where this sequence is hypothesized, we ordinarily want to know how "important" the political process variables are as compared to the less "political," background factors. There is a severe difficulty in making this evaluation, however. Not only do we have multicollinearity, but our ability to measure these variables adequately leads us into a very serious form of this

problem. For example, our socio-economic variables come much closer to being continuous variables, while the process measures usually are much more limited ordinal indicators at best. We therefore have a strong chance at measuring much greater variance in the background than in the process variables. If we add this additional condition, differential variance measured, to the usual specification of multicollinearity, we can derive the conclusion that we are likely to find the process variables to be statistically insignificant and the background variable to do all of the "explaining" even though the causal sequence pictured above is very close to the truth.

First let us define a constant " κ " as the ratio of the variances in two independent variables. Thus

$$\kappa = \frac{\sum x_2^2}{\sum x_1^2} \quad 1.)$$

Then we can show that as κ increases (the variance in x_2 becomes much greater than the variance in x_1) we are very likely to have

$$V(\hat{\beta}_1) > V(\hat{\beta}_2) \quad 2.)$$

That is our β_1 estimate is more likely to fail the test of significance than our β_2 estimate, even though the following model is the "truth."

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \epsilon_1 \\ x_1 &= \beta_3 x_2 + \epsilon_2 \end{aligned} \quad 3.)$$

Applying the ordinary estimating procedures to 3. gives the conventional variances of the β estimates (notice that in applying these to 3. we have controlled x_1 for the effects of x_2 and vice versa.) Thus

$$V(\hat{\beta}_1) = \sigma_{\epsilon_1}^2 / \sum x_1^2 (1 - r_{x_1 x_2}^2) \quad 4.)$$

and

$$V(\hat{\beta}_2) = \sigma_{\epsilon_1}^2 / \sum x_2^2 (1 - r_{x_1 x_2}^2) \quad 5.)$$

As the ratio of the variance of the β_1 to the β_2 estimates increases, the more likely will x_1 appear less significant than x_2 . If we substitute into that ratio the right hand sides of 4 and 5, we have

$$\frac{\sigma_{\epsilon_1}^2 / \sum x_1^2 (1 - r_{x_1 x_2}^2)}{\sigma_{\epsilon_1}^2 / \sum x_2^2 (1 - r_{x_1 x_2}^2)} = \frac{\sum x_2^2}{\sum x_1^2} \quad (6.)$$

Therefore as the ratio of the two variances (κ) increases, the more likely are we to find differential significance between the two variables.

APPENDIX D

INSTRUMENTAL VARIABLE APPROACH TO MULTICOLLINEARITY

Instrumental Variable Approach to Multicollinearity

John Aldrich

The increasingly common use of regression and related techniques in social science has highlighted a series of statistical and inferential problems. Undoubtedly the most frequently cited of these is multicollinearity; the existence of linear relationships among independent variables. Unfortunately, while the problem of multicollinearity is well known and while there are a number of excellent analyses of the limitations it poses and several methods of demonstrating its existence and extent, there are relatively few useful procedures to avoid and/or alleviate its consequences. Most analysts conclude that the researcher should gather new data to break the multicollinear pattern and/or incorporate new information into the analysis by assumption, theoretical deduction or previous analyses. The first strategy of gathering new data is obviously infeasible in most circumstances. Moreover, we are rarely in a situation in which theory, past results, and the like are of sufficient strength to support the second strategy.

The purpose of this paper is to consider a procedure to ameliorate inferential problems posed by multicollinearity in one of the more common situations in social research; a situation in which multicollinearity poses some of its most severe interpretative and inferential limitations. This situation (or, more properly, set of situations) is that in which the researcher has a relatively large number of independent variables. In such cases, the researcher is often not interested in the effects of each individual variable, per se. Rather, the question focuses on the effects of sets of variables (where a set may consist of only one variable but often consists of many variables). For example, independent variables may be grouped into general types such as economic, political, and social variables or, as in Coleman, et al's research (1966) on school achievement and background characteristics. Alternatively, the researcher may have a number of control variables in addition to his variables of specific interest, or he or she may have exogenous and endogenous independent variables. A final example would be the situation where the

researcher has multiple measures of one or more variables (e.g., education, occupation and income as multiple indicators of social status or, as in our example later in this paper, more specific and direct multiple indicators). The procedure to be developed in this paper is applicable to this type of situation. The existence of multicollinearity can pose quite complex difficulties in such circumstances due to what Gordon (1968) refers to as "repetition" and "redundancy".

This paper, therefore, will focus on the development and assessment of a very practical technique to deal with the specific problem outlined above. Applied statistical techniques often become far removed from the statistical model on which they rest and hence run the risk of being applied inappropriately. Therefore, we will attempt to touch base with the statistical model and its assumptions as frequently as possible.

In the next section we will quickly review the OLS model and the properties of its estimators. The nature of multicollinearity and its consequences will also be raised. Succeeding sections will state the type of situation for which our procedure is intended, develop the procedure itself, assess its properties, and address the kinds of questions and hypotheses for which the technique is suitable. We will then provide some example applications and a concluding discussion.

OLS and Multicollinearity

We assume the basic OLS model. That is, we assume that there is a correctly specified, single equation model with Y as the dependent variable and the primary focus of the research. Y is assumed to be linearly and additively related to a series of k independent or explanatory variables X_j , $j=1,2, \dots, k$, and a stochastic

term U . For simplicity but without loss of generality, we also assume that all variables are measured in deviations about their mean, thereby enabling us to ignore the intercept or constant. Therefore, the basic model is that the i^{th} value of Y is equal to a linear combination of X_j 's and U :

1. a. $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$ or in matrix notation
- b. $Y_i = X_i \beta + U_i$

We wish to estimate the unknown parameters, β_j . To do so, we collect a sample of n independent random vectors of observation on Y and X_j . We also make the following Gauss-Markov assumptions about the nature of the stochastic term, U :

2. a. $E(U_i) = 0$
- b. $E(U_i^2) = \sigma^2$
- c. $E(U_i U_j) = 0 \quad i \neq j$
- d. $E(U_i X_{ji}) = 0 \quad \text{if } X_{ji} \text{ are random variates}$

Combining these assumptions with our sample data, the OLS estimators, $\hat{\beta}$, of the parameter vector β may be obtained by the following equation:

$$3. \quad \hat{\beta} = (X' X)^{-1} X' Y$$

The estimators have the sampling distribution:

$$4. \quad \hat{\beta} \sim N(\beta, \sigma^2 (X' X)^{-1})$$

The OLS assumptions "justify" the estimators, $\hat{\beta}$, in the sense that a.) they are unbiased estimators, b.) of linear, unbiased estimators, they are statistically best (or are BLUE) since they have the smallest possible sampling variances in their class and hence are most precise, and c.) we know that the estimators are normally distributed with mean and variance as given in eq. 4, and we can therefore perform various statistical tests.

At the same time that these assumptions allow us to reach the above conclusions, they also limit the range of applicability of the model. For example, the assumption that the stochastic term has zero expectation must be plausible. One implication

of this assumption is that the model under test is properly specified. Thus, we are assuming throughout the paper (and whenever OLS is employed) that we have included all relevant or important variables and none that are irrelevant, that we have included these variables in the proper form (i.e., that linearity and additivity in the parameters is the correct mathematical form), and that any measurement error in the independent variables is small. This list of implications of the assumptions in using OLS can be extended easily. However, even a fully exhaustive list would show no implication concerning anything less than complete (perfect) multicollinearity. That is, anything short of the pathological degeneracy of complete multicollinearity presents no statistical problems whatsoever. To see why this conclusion holds and yet why multicollinearity presents problems of inference and interpretation, let us review the nature of multicollinearity.

Multicollinearity, as its name suggests, concerns the existence of linear relationships between a series of independent variables. The existence and extent of multicollinearity among the X_j 's may be wholly unrelated to the nature of the relationship between these independent variables and Y , the dependent variable. The first point which should be kept firmly in mind is that the question is not whether multicollinearity is present or absent but the degree to which the X matrix is multicollinear. Rarely will we ever find complete multicollinearity, and rarely will we find it completely absent. Rather there will always be some multicollinearity, and the question is how much. Second, a high pairwise correlation between two independent variables, a relation easily discovered, is sufficient to guarantee a high degree of multicollinearity. Large correlations are not, however, necessary. With several independent variables there may be relatively small pairwise linear relationships, and there may still be a high degree of multicollinearity. As a final general point, serious multicollinearity might exist only among a subset of independent variables. However, the consequences of such a pattern may well affect the inferences and interpretations made about all

independent variables

We stated earlier that anything short of complete or perfect multicollinearity presents no statistical problems, since it violates none of the assumptions of OLS. Perfect multicollinearity means that there is complete linear dependency within the set of independent variables (e.g., X_j and X_k may have a correlation of ± 1). If so, OLS estimates are undefined. Short of this unusual occurrence, even a very high degree of multicollinearity violates no OLS assumption. Therefore, all deductions from those assumptions (such as those described above) continue to hold. OLS estimators continue to be the best linear and unbiased estimators it is possible to obtain, given that model specification (which is assumed to be the correct one) and that data set. Moreover, the distribution of $\hat{\beta}$ remains that defined in equation 4. Therefore, from a purely statistical point of view, we can reach all of the same conclusions as we could if there were no multicollinearity. Moreover, regardless of whatever consequences multicollinearity might have, less than perfect multicollinearity has no effect whatsoever on any of the "overall" measures. Thus, R^2 , the "overall" F value, the estimate of the stochastic variance, and the prediction of Y by $\hat{Y}$ are totally immune to multicollinearity.

The conclusion that even very high degrees of multicollinearity present no problems, statistically, is only one face of the usual schizophrenic attitude towards multicollinearity. The other face is that it has serious consequences, especially considering that there appears to be so little one can do to alleviate the situation. The most important consequence of multicollinearity is that the precision of individual estimates falls. While the estimates are still the best--the most precise possible given the model and data set--"best" may not be very good. In particular, the sampling variances of $\hat{\beta}$ will be larger than if the independent variables were independent. Johnston exemplifies this point (1972, pp. 160-161). In a simple model with two independent variables, he shows that

increasing the correlation between the two variables from .5 to .9 increase their sampling variances by 300%, while increasing the correlation from .5 to .95 increase the sampling variance 750%, everything else remaining constant. The assumptions he makes leads to the following formula for the expression of the variance of both $\hat{\beta}_j$, where σ^2 denotes the (constant) stochastic variance and r^2 the squared correlation between the independent variables (Johnston, 1972, formula 5-84, p. 161):

$$5. \text{ var } (\hat{\beta}_j) = \frac{\sigma^2}{(1-r^2)}$$

The relationship between the degree of collinearity (r) and the variance of both of the two parameter estimates is plotted in figure 1 for $\sigma^2 = 1$, and positive collinearity. As can be seen, only very high correlations lead to significant inflation of the variances (and hence standard errors) of the estimates. However, the inflation, when it does begin to grow, takes off very quickly and literally grows without bound.

-----FIGURE 1 ABOUT HERE-----

Due to the interdependencies among the independent variables, there will also be a covariation (usually negative) between the sampling distributions of the estimators. As with their variances, the covariance will also increase sharply (away from zero) with increasing multicollinearity. While the expectation of each $\hat{\beta}_j$ will still be β_j , if one parameter estimate is, by chance, too large, then some other estimate(s) will likely be underestimated. Johnston provides the following formula for the covariance between the sampling distributions of the two parameter estimates for his two variable case (1972, eq. 5-85, p. 161):

$$6. \text{ cov } (\hat{\beta}_1, \hat{\beta}_2) = \frac{(-r) \sigma^2}{(1-r^2)}$$

With $\sigma^2 = 1$ and a positive correlation, the relationship between the correlation between the two independent variables and the covariance between their two parameter

estimates is plotted in Figure 2. This relationship is the mirror image of that in Figure 1. When the two variables are uncorrelated, there is little covariation between the estimates, and when there is a small or even moderate correlation, the covariance is relatively small. However, as the collinearity becomes severe, the covariance increases rapidly and without bound.

[FIGURE 2 ABOUT HERE]

The combination of large standard errors of estimates with large covariances implies that it becomes very difficult to separate out the individual effects of the various independent variables and their relative influence. Quite often, one of the most important questions of the research is precisely that--to assess which subsets of variables are the most important determinants of the dependent variable. It remains true that there is no loss of precision in predicting Y from the full set of X_j 's. The analysis of any individual variable or subset of variables is, however, seriously affected. Moreover, as Gordon notes by some quite clear and remarkable examples (1968, pp. 592-616), the estimates may be quite sensitive to even small variations in the interrelationships among the independent variables. He also shows that even if the multicollinearity is concentrated in a subset of independent variables, all coefficients may be affected, even those for relatively independent variables. One immediate consequence is that the estimates obtained are likely to be quite sensitive to peculiarities in the particular sample data employed.

These three problems - larger than desired standard errors of estimates, relatively high covariances, and sensitivity of the estimates to small perturbations - lead to two quite serious consequences. First, tests of significance on individual coefficients or subsets of coefficients may show that they are not significantly different than zero or other null value when the true state is

that they are. That is, we may conclude that a variable (or subset of variables) has no effect, not because it "really" has no effect, but because our sample data have not enabled us to discern the significant true impact. Second, the researcher may then erroneously conclude that a variable should be dropped from the analysis as "irrelevant." If so, the model will then be underspecified and therefore estimates of the "reduced" model will no longer be best or unbiased (or consistent). To the extent that empirical estimation is used as one element of the decision concerning model specification, one effect of multicollinearity may be to underspecify models. To the extent that this holds, the conclusion that multicollinearity presents no statistical problem no longer applies for this chain of reasoning leads us to conclude that biased estimates may result due in part to multicollinearity.

There are a whole range of additional problems concerning the interpretation of the estimated equations that may be traced to even moderate amounts of multicollinearity. Two examples may make the point most clearly. Gordon notes that one interpretation of an OLS coefficient is that it represents the influence of an independent variable on the dependent variable controlling for all other independent variables included in the estimation. In a section he entitles "The Partialling Fallacy," he says (1968, p. 593):

The object, of course, is not simply to destroy an observed relationship but, rather, to see whether it can be destroyed by controlling for a variable that has been hypothesized to be potentially relevant and conceptually distinct Consequently, variables are often introduced as controls that are not meaningfully different in terms of what would constitute an appropriate theory. So closely do these variables approach being identical with one of the variables already in the zero-order relationship that controlling for them becomes tantamount to partialling the relationship out of itself.

and (p. 594 and p. 595):

. . . there is no statistical rule for attributing controlled covariation to the influence of one rather than another of the independent variables, regardless of the disparity in size between their partial correlations. . . . There is nothing

more fundamental about a partial, as compared to a zero-order association, unless a good theory makes it so.

Thus, the indiscriminant application of controls can be quite misleading, and misleading in at least two ways. The first is that with several measures of a common concept, controlling any one of the measures for any one of the others "partials out" valid variation. Noell, in a communication (1970, pp. 1249-1250) discusses an article that attempts to account for redistribution by socio-economic and political variables (Fry and Winters, 1970, pp.508-522). Noting for example, that the zero-order correlation between median income and redistribution is .18, but when partialled it is -.27, Noell states (1970, p. 1249, emphasis in original):

. . . The independent variables suffer from multicollinearity . . ., so that when one "controls" for variation caused by, say, industrialization, urbanization and education to see if "median income" has any "independent" effect on "redistribution ratios," one actually controls the valid variation of, in this case, "median income."

The second point is that "there is no statistical rule for attributing controlled covariation to the influence of one rather than another of the independent variables" (Gordon, 1968, p. 594) or, of course, one set of independent variables rather than another. Such attribution must be justified theoretically and receives no support statistically. Much of the debate about the relative importance of school and background factors in affecting students' achievements that surrounds the "Coleman report" can be traced to the simple fact that these sets of independent variables are interrelated.

Existing Procedures Concerning Multicollinearity

In the last section, we concluded that while multicollinearity presents no statistical problems, it leads to numerous inferential difficulties. Further, statistical properties of OLS estimates may be lost (and with them, the justification for using OLS) when using regression as a guide for model specification.

This paper proposes several procedures to minimize these inferential and interpretative problems in some common situations. Before venturing into new territory, however, the existing state of the art should be reviewed.

The first and most obvious question concerns the demonstration of the presence of multicollinearity. The simplest approach is to scan the pair-wise the matrix of correlation between independent variables. A large correlation does indicate a high degree of collinearity and is therefore a sufficient indication of serious multicollinearity. Unfortunately, large zero-order correlations are not necessary. For example, X_i may be correlated with seven other X_j 's at, say, .15 each, and X_i may be a perfect linear combination of the other independent variables. In fact, the largest correlation sets the lowest bound on the extent of multicollinearity. Therefore, if the correlation is statistically significant, the hypothesis of no multicollinearity may be related. However, if the correlation is not significant, no conclusion may be drawn.

A second procedure builds off the most direct consequence of multicollinearity; the inflated standard errors of the estimates. Since the estimate of the stochastic variance, $\hat{\sigma}^2$, is unaffected by multicollinearity, the formula for the sampling variances of the estimates indicates that increasing interrelationships in X , the data matrix, will immediately affect these variances:

$$7. \text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

Therefore, standard errors of estimates are adequate indicators. Based on the same root cause, increasing multicollinearity can be seen directly from the determinant of $(X'X)$, which will approach zero, or its inverse $(X'X)^{-1}$, which will approach infinity as multicollinearity increases. These three figures are often readily obtainable, but are difficult to interpret without some basis of comparison.

One approach to demonstrating the existence of the problem may also be used to assess the magnitude of the problem, and further, has been proposed as a remedy

(these last two questions will be taken up shortly). This procedure is to perform a principal components analysis of the full set of k independent variables (see Johnston, 1972, pp. 322-331). This technique transforms the k variables into a new set of k variables that are orthogonal or independent. Further, the first transformed variable is that linear combination of X_j , which has the largest percentage of the total variance in the set of independent variables (i.e., is the most collinear combination). Incidentally, if the last or k th component has zero sum of squares, there is complete multicollinearity in the data set. Bartlett's sphericity test (Cooley and Lohnes, 1971, pp. 102-114) has the null hypothesis that the k original variables are independent in the population. Rejection of the null implies that there is a statistically significant degree of multicollinearity. This test is suitable for small samples and is clearly more powerful than tests based on individual correlations.

Another "existence" test has been proposed by Farrer and Glauber (1967, pp. 92-107). They propose examining the linear relationship between each independent variable and all others. From such an examination, one can obtain F statistics. The largest F will isolate the single variable most highly collinear with the set of other independent variables, and may be used to test for a statistically significant departure from independence.

These tests for significance of multicollinearity are probably rather uninteresting. Most situations, especially those with large samples and several to many independent variables, will likely result in the rejection of the hypothesis of no interdependency. Of more interest are attempts to isolate the most severe patterns of multicollinearity and to measure the degree of it, rather than just its existence.

Several of the aforementioned procedures address these two questions. For example, the Farrer and Glauber procedure is equivalent to regressing each independent variable in turn on all of the $k-1$ remaining independent variables. The

coefficient of multiple determination, R^2 , indicates the extent to which each variable is interrelated with the others (and the associated F statistics will be obtained for tests of significant interrelationships). Moreover, the largest R^2 indicates which independent variable is most seriously affected by multicollinearity. The coefficient estimate also provide an indication of the particular linear combination of the $k-1$ variables that is most collinear with the k th variable. The principal component analysis may also be used to isolate and measure the degree of multicollinearity. Recall that the first of the k components is precisely that linear combination of the measured variables with the largest shared variations in the sample data. Most principal component analysis (or factor analysis) packages report the sum of squares associated with each component. The sum of these k sums is the total sum of squares in the independent variables, and the ratio of the sum of squares of the first component to the total is a direct measure of the extent of multicollinearity.

The Farrer and Glauber procedure allows one to isolate the particular variable most seriously affected by multicollinearity. It does not, however, provide a very precise indication of the particular configuration of new data (assuming them available) that would most efficiently reduce multicollinearity. The principle component analysis indicates the configuration of the independent variables most collinear through the variable "loadings" on the first component. Gathering new data of a pattern orthogonal to those loadings would reduce the collinearity along this component (and presumably one could design a configuration of data values which would not significantly raise the multicollinearity along the second, third, etc. component). Both of these approaches assume that the researcher desires to minimize the interrelations in the independent variable data matrix. Silvey (1969, pp. 539-552) has proposed a technique with a slightly different objective. Instead of directly attacking the $(X'X)$ matrix, he focuses his attention on the most direct consequence of multicollinearity; in-

lated standard errors of estimates. His procedure is designed to answer the question: "What additional data would most effectively reduce the size of any specific standard error." He answers the question through a variation of principal components. Let V be the matrix of "latent vectors" associated with the principal component analysis of $(X'X)$, and let c be a known linear combination of the $\hat{\beta}$ coefficients. For example, if C is a vector of k 1's, $C' \hat{\beta} = \sum_{j=1}^k \hat{\beta}_j$. C can be expressed as a linear combination of V , with the solution Silver proposes:

$$8. (I + X'X)^{-1}C$$

Farrar and Glauber critique a "rule of thumb" procedure for deciding if the extent of multicollinearity is "acceptable" or "harmful." They quote Klein's version of the rule (1962, p. 101) as stating that "intercorrelation or multicollinearity is not necessarily a problem unless it is high relative to the overall degree of multiple correlation." (1967, p. 98). While this is simply a guide, the rule does suggest that "moderate" levels of multicollinearity may be "harmful" and that what is important is the relationship between the independent variable as they are related to the dependent variable.

The above procedures isolate, measure, and/or test the significance of the multicollinearity in the data set. Let us now consider several procedures that have been proposed as solutions to the problem.

Two proposed remedies may be dealt with quickly. The first and most frequently encountered suggestion is to gather new data. As the previous review has indicated, many diagnostic techniques have been designed with the purpose of assessing just what configuration of data would be most useful. Such controlled data collection possibilities are limited. Of course, increasing the size of the data set, per se, may help, since standard errors are themselves functions of, inter alia, sample size. Gathering more data is often only slightly less unrealistic than gathering more data of a particular type. A second approach is to in-

corporate either a priori knowledge (say, estimate of specific parameters based on previous research) or theoretical assumptions or deductions. The usual approach is to incorporate this non-sample based information through constrained OLS or some generalization (for details see Johnston, 1972, pp. 164-165, 221-227). As Johnston puts it (1972, p. 164) additional information (new data, otherwise estimated relationships or theoretical assumptions/deductions) may be called for since, with the inadequate data at hand, the researcher can not "make bricks without straw."

A series of proposals recommend one form or another of orthogonalizing the independent variables. The most obvious orthogonalization procedure is to use the results of a principal components or factor analysis (with, perhaps, a varimax or simple structure rotation) of the independent data matrix. This approach was apparently first suggested in this context by Kendall (1957, see also Gorsuch, 1973, and Stone, 1945 for elaboration and application). The difficulties with this approach are the well known difficulties of factor analysis in general. In particular, the resultant orthogonal variables may not be substantively interpretable (for further evaluation of the strengths and weaknesses of this approach, see Farrer and Glauber, 1967, pp. 96-98).

Gorsuch (1973) outlines another series of techniques that have the same objective of yielding orthogonal independent variables. The problem, he says, is that the two (or more) independent variables overlap. Therefore, he considers various procedures for dividing this overlapping variance. "First," he says (pp. 90,91), "the overlapping variance can be attributed to one of the two variables." This is essentially the procedure Coleman et al, followed in their analysis of school achievement. They deal with school factors and home and related background factors, and "attributed the overlap" to the non-school factors. "Second, overlapping variance can be divided between the two variables." Many division rules can be devised, the most common being that employed whenever one uses multiple regression.

"Third, the overlapping variance can be represented by a new variable," derived from the two observed independent variables. In this procedure, the effect of the original variables is only that due to their non-overlapping components or their independent effects. In general, there may be many derived variables, each representing sections of the overlap between each pair, triad, etc. of observed variables. The procedure Gorsuch suggests for deriving these new variables is "image analysis," a close relation to factor analysis. There is no statistical criterion for choosing a "best" method of dividing shared or overlapping variance. This choice can only be based on theoretical grounds. For example, the authors of "Coleman Report" make a theoretical argument that their particular statistical procedure is substantively justifiable. Much the same type of theoretical argument is made in a study of the effects of zoning on land prices (Maser, Riker, Rosset, 1974). Their argument is, essentially, that the political act of imposing zoning restrictions only has a modifying effect on economic market-place factors and that, therefore, any overlapping effects of zoning and the market on land prices should be attributed to the causally prior economic forces (Gorsuch's first procedure). Note that their argument must be evaluated on purely non-statistical grounds, even though it has statistical implications, for it is a theoretical argument.

A final proposed technique for handling multicollinearity is known as "ridge regression." OLS estimators are the most efficient possible estimators within the class of linear and unbiased estimators. It is possible, however, that a more efficient estimator exists within the class of biased estimators. Ridge regression is a technique that purposefully introduces bias in the estimators of the parameters to increase efficiency and, more importantly, to produce more "precise" estimators, i.e., estimators that produce a smaller mean squared error and thus attempt to minimize error considering both the variance of the estimator and the bias introduced. Unfortunately, ridge regression does not appear to be

applicable for the sorts of situations to be considered in this paper. First, any real gain over OLS is generally obtained only when there is very severe multicollinearity (e.g., a correlation between two independent variables exceeding .9). Secondly, the procedure is practical only when there are a small number of independent variables (say, two). Finally, the estimating procedure is complex, iterative (and hence yielding no determinant solution), and non-linear, and thus is impractical for the sort of situation analyzed here.

An Instrumental Variables Approach to Multicollinearity

The procedure we propose is designed for a specific set of circumstances. In particular, we assume that there are a large number of independent variables. This is a fairly common situation in much of social science research and policy analysis. For example, the applications reported below, based on the "Coleman data" have ~~more than 70~~ variables. To be sure, such a number is extreme. Yet there are many circumstances in which there are more than just three or four or even ten independent variables. These situations where there are many independent variables often share other characteristics. First, the independent variables are often moderately to severely multicollinear, and even with only moderate amounts of multicollinearity there can be severe interpretative and inferential problems. Recall the "rule of thumb" which says that the degree of multicollinearity may be considered severe if it is high relative to the multiple correlation between the dependent and full set of independent variables. In situations where there are large numbers of independent variables, this boundary is often surpassed. Moreover, this situation is exactly that that is most likely to fall victim to the problems of "repetitiveness" and "redundancy" dramatically illustrated by Gordon (1968). Finally, and most importantly here, research under these conditions is rarely focused on the effect of each individual variable. Rather, many of the variables are classifiable into sets or types of variables. For example, there may be a few individual variables of direct concern, the remaining being considered only as "controls." Alternatively, there may be a set (of one or more) "economic," "political," and "social" variables or some other such set of groupings. Similarly, there may be several measures of the same phenomenon, and the interest is in the impact of the common phenomenon and not each individual measure. Whatever may be the particulars

of the specific situation, some such simplifying procedure is invariably employed when the researcher has a large number of independent variables. Moreover, we should emphasize that data reduction procedures (e.g., the principle components approach) lose contact with individual variables. Our procedure does not.

The assumptions of the technique are as follows. First, it is assumed that the regression equation including the full set of individual variables meets the assumptions of ordinary least squares. In particular, the single equation model is correctly specified (i.e., the set of independent variables includes all relevant and no theoretically irrelevant variable and is linear in the parameters) and, in addition, the stochastic term satisfies the Gauss-Markov assumptions. Thus, there are k explanatory variables, X_j ; $j = 1, 2, \dots, k$. Next, we assume that there is a partition of X_j into p exclusive and exhaustive sets, Z_m , $m = 1, 2, \dots, p$; $p \leq k$, such that each X_j falls into one and only one Z_m . Note that if $p = k$, then each variable forms its own, single-element set. If $p < k$, then at least one set in the partition contains at least two original variables. In general, $p \ll k$, nonetheless any given Z_m might contain but a single X_j . This assumption, then, is the only additional assumption which must be made, and we suggest that, with large j , it usually is made in practice.

One way of looking at the problem of multicollinearity is that there is too little information supplied by the substantive theory in specifying the equation and the sample data set in estimating the parameters (what Johnston refers to as "making bricks without straw"). As we have seen, most remedies propose adding new information to the estimation process, either by gathering new data or by incorporating theoretical assumptions or deductions or previous results. In this case, the new information is supplied theoretically by

defining the partition of X_j into a presumably much smaller set of Z_m . As with any other assumption, it must be justified by the researcher on a priori grounds and on a case-by-case basis. Further, as with any other assumption, it will apply only approximately in any actual situation, which is only to say that the plausibility of the assumption will vary. Nonetheless, of all possible theoretical assumptions which can be used to provide a lever on inferential problems caused by multicollinearity, it is very mild and is often implicitly or explicitly employed.

These assumptions state that the fully specified population equation (recalling that we have arbitrarily but without loss measured X_j in deviations from their mean to ignore the intercept term) is:

$$9a) \quad Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i \quad \text{for } i = 1, \dots, n$$

observations, or

$$b) \quad Y_i = X_i \beta + U_i$$

The partitioning assumption means that each X_j is an element of one and only one Z_m . Define a one-to-one correspondence of each Z_m to a V_m such that:

$$10) \quad \begin{aligned} V_{1i} &\triangleq \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_r X_{ri} \quad \text{where } X_1, \dots, X_r \in Z_1 \\ V_{2i} &\triangleq \beta_{r+1} X_{r+1,i} + \dots + \beta_s X_{si} \quad \text{where } X_{r+1}, \dots, X_s \in Z_2 \\ &\vdots \\ V_{pi} &\triangleq \beta_{t+1} X_{t+1,i} + \dots + \beta_k X_{ki} \quad \text{where } X_{t+1}, \dots, X_k \in Z_p \end{aligned}$$

Then, equation 9 may be written equivalently as:

$$11a) \quad Y_i = V_{1i} + V_{2i} + \dots + V_{pi} + U_i \quad \text{or}$$

$$b) \quad Y_i = V_i + U_i$$

Since Z_m and therefore V_m is presumably much smaller than X_j , a great deal of reduction has been effected, but a reduction that focuses on the partition of variables that is theoretically of interest.

To estimate the parameters, consider the following two step procedure.

First, estimate equation 9 by OLS:

$$12) \quad Y = X \hat{\beta} + E$$

This estimation yields statistically useful and appropriate estimators, i.e., yield BLUE estimators with known distributions, since multicollinearity does not bear upon the statistical properties of the estimators. This conclusion provides the statistical justification for proceeding. Next, construct the "instrumental variables," the $\hat{V}_m$:

$$13. \begin{aligned} \hat{V}_1 &\triangleq \hat{\beta}_1 X_1 + \dots + \hat{\beta}_r X_r \quad \text{for all } X_j \in Z_1 \\ \hat{V}_2 &\triangleq \hat{\beta}_{r+1} X_{r+1} + \dots + \hat{\beta}_s X_s \quad \text{for all } X_j \in Z_2 \\ &\vdots \\ \hat{V}_p &\triangleq \hat{\beta}_{t+1} X_{t+1} + \dots + \hat{\beta}_k X_k \quad \text{for all } X_j \in Z_p \end{aligned}$$

The second step consists of an OLS regression of Y on V_m :

$$14a) \quad Y_1 = \hat{\alpha}_1 \hat{V}_{11} + \dots + \hat{\alpha}_m \hat{V}_{m1} + e_1 \quad \text{or} \\ b) \quad Y_1 = \hat{V}_1 \hat{\Gamma} + E_1$$

Given that equation 9 satisfies the stated assumptions and Z_m and hence V_m form the theoretically appropriate partition, then the instrumental variable $\hat{V}_m$ and associated estimators, $\hat{\Gamma}$, have all the properties of any appropriate instrumental variable (as shown in most econometric texts, e.g., for proof of consistency and related derivations, see Johnston, 1972, pp. 278-281). If $m=k$, equation 14 is identical to equation 12. Therefore, the degree to which k exceeds m is a measure of the additional theoretical leverage supplied.

We will consider particular conclusions that can be drawn from using this procedure shortly. Let us first note precisely what effect this has had on the multicollinearity problems. First, we have significantly reduced the number of parameters to be estimated and interpreted. Thus, we are far less likely to conclude that the partition has no effect on the dependent variable than any or all of its component parts due to the inordinately large sampling variances of the $\hat{\beta}_j$ caused by the multicollinearity. Moreover, the consequences of multicollinearity attributable to variables falling within any one member of

the partition (i.e., a Z_m) is completely eliminated in estimating equation 14. This result is particularly crucial since variables which fall into the same class theoretically are the most likely to be jointly collinear. The clearest case is where the elements of any member of the partition are alternative measures of the same underlying phenomenon. These multiple measures are expected to be very highly interrelated, far more so than variables across Z_m 's. In general, if the theoretical partition is meaningful, then the most severe instances of multicollinearity will be internal to each given member of the partition. This multicollinearity is entirely removed in estimating the second step, equation 14. Note that the suggested procedure does not entirely eliminate multicollinearity. Rather, it eliminates what is expected to be the most severe components of it. What is left, of course, is any multicollinearity between theoretically distinct types of independent variables.

Second, there is some virtue to the fact that the number of instrumental variables in eq. 14 is usually much smaller than the original set of independent variables. If the remaining cross-partition multicollinearity is suspected to be important, then isolating it, measuring its extent, and/or using some other procedure (e.g., ridge regression) to alleviate it will be greatly simplified.

Third, some of the most severe interpretative and inferential problems due to multicollinearity are attributable to what Gordon calls "redundancy" which he defines (1968, p. 596) as, "high correlation between two or more independent variables," and "repetitiveness," which he defines (1968, p. 596) as, "the number of redundant independent variables, regardless of the degree of redundancy." The consequences of redundancy are obvious. Concerning repetitiveness

he says (1968, p. 595, emphasis here, as above, in the original):

The argument developed above helps us to realize that among the independent variables there could occur two or more subsets of variables, the members of which were redundant (strongly correlated) with variables in their own subsets but relatively independent (weakly correlated) with respect to variables in other subsets. It becomes immediately apparent that, under these circumstances, the relative size of a variable's regression coefficient depends to a considerable extent upon the number of other variables in its subset.

The use of this procedure completely eliminates the theoretically predicted repetitiveness and redundancy.

Hypothesis testing is straightforward. As noted earlier, coefficient estimates for instrumental variables, such as $\hat{\eta}_j$, are consistent estimators with asymptotically normal sampling distributions. The fact that these properties are for large samples is of no concern, since the procedure, presupposing large numbers of independent variables, would rarely be applied to small samples. Therefore, to test the hypothesis that an $\hat{\alpha}_m$ is different from zero (or any other null hypothesis) one merely uses the Z-score test. The t-or F-test usually reported in regression output is marginally more conservative, converging to the Z test in large samples. Such a hypothesis is of the form "do school variables exert a significant effect on school achievement," using the Coleman analysis for example. Tests concerning several parameters may be performed as in O.L.S. regression, and, of course, the "overall significance" is the usual F-test. The latter measure, along with the R^2 , is unaffected by multicollinearity and is therefore testable from the original OLS regression, equation 12. The same procedure holds for predicting Y_i by $\hat{Y}_i$, as well as examining any other characteristics of the total equation.

The "relative importance" of the instrumental variables may be assessed in the usual fashion. For example, the standardized betas of each $\hat{\alpha}_m$, the

correlations between Y and individual $\hat{V}_m$, partial correlations, etc., may be investigated subject to the usual (and, unfortunately, often debilitating) caveats.

Notice that, given equation 13, $\hat{\alpha}_m$ is constrained to be one (providing a useful check on what is often a tedious and error-filled step in solving the identities of eq. 13). Therefore, attention is focused directly on the sampling distributions of the estimates. Notice, too, that $\hat{V}_m$ is a weighted linear combination of the theoretically related variables, members of the same partition. The weights are best and unbiased estimates of the true weights. Therefore, $\hat{V}_m$ represents that estimated linear combination of the related independent variables which is most strongly related to, or best predicts, the dependent variable. Further, a one unit change in $\hat{V}_m$ represents that change in the within partition X_j 's which predicts the largest change in Y .

In general, the most severe multicollinearity will be eliminated through this procedure. However, there will likely remain cross-partition multicollinearity. An important question, then, is the extent to which the relationship between a $\hat{V}_m$ and Y is affected by remaining multicollinearity. In many respects, this question is more important than the degree of multicollinearity, *per se*.

For example, one might ask of the Coleman data whether school variables significantly reduce the impact of background variables on achievement due to the collinearity between these two sets of variables. To judge the effects of covariation between, say, $\hat{V}_1$ and $\hat{V}_2$ on the assessed influence of each, first form the instrumental estimators:

$$14) \quad Y = \hat{\alpha}_1 \hat{V}_1 + \hat{\alpha}_2 \hat{V}_2$$

Compare these results to the following regression of Y on $\hat{V}_1$ and $\hat{V}_2$ individually:

$$15) \quad Y = \hat{\zeta}_1 \hat{V}_1$$

$$\text{and} \quad Y = \hat{\zeta}_2 \hat{V}_2$$

The hypothesis, then, is that $\hat{\delta}_1 = \hat{\alpha}_1$ (or $\hat{\delta}_1 = 1$) versus $\hat{\delta}_1 > \hat{\alpha}_1$ (since, in general, ignoring a relevant partition will inflate the impact of included partitions). Rejection of the null of equality allows one to infer that the effect of the partition is significantly affected by (multi-) collinearity. The degree of this effect is directly measured by the difference between $\hat{\delta}_1$ and $\hat{\alpha}_1$. That is, the difference represents the extent to which the effect of the partition on Y is reduced due to cross-partition collinearity having removed all within partition multicollinearity.

This measure of the consequence of multicollinearity is based on the fact that $\hat{V}_m$ is already a consistent estimator the true effect of the instrument. This measure is therefore different from a purposefully misspecified estimator. For example, consider the result of regressing Y on just the members of one partition in a misspecified form of eq. 12 and then constructing the biased instrument, say $\hat{W}_m$, based on solving the parallel to eq. 13. The comparison, then, would be between regressing Y on the full set of biased estimators to regressing Y on each biased estimator individually. In this case, the coefficient estimate for regressing Y on an individual $\hat{W}_m$ would be, identically, one. The coefficient estimates from regressing Y on the full set of $\hat{W}_m$'s would indicate the degree to which the misspecified model overestimates the effect of a single set (or set of sets) of independent variables. Such an estimator might be appropriate if the investigator disagreed with the original research which posited a definite sequential ordering of independent variables and thereby attributed all overlap due to multicollinearity to the assumed prior variables (as in the Coleman and Maser, et al examples cited above). At least the misspecified estimation would enable the investigator to assess the importance of the causal ordering initially assumed.

Further hypotheses may be entertained as the research concerns dictate. Let this section be closed by emphasizing that the a priori most severe sources of multicollinearity are entirely removed by this procedure, that the estimators have known and useful properties, and that many important research questions can be answered through this procedure. Information reduction procedures are often criticised by saying that one loses precious information in the reduction (e.g., factor analysis often leaves one with uninterpretable independent variables). The procedure suggested here focuses on the theoretically most important groupings of variables. Moreover, the first step in the procedure supplies the best linear unbiased influences of each individual variable prior to summarizing. Therefore, no information is irretrievably lost.

An Example Application

To exemplify the utility of the proposed procedure, we will apply it to the data analyzed in the Report on Equality of Education Opportunity (Coleman, et al, 1966). The impetus behind this example and indeed behind this project, is a major re-analysis of these data being conducted by Charles Cnudde (see, in particular, Cnudde, 1974, 1976. I am very grateful for his kind assistance and his generous permission to report some of his findings.)

This analysis seems to be a particularly appropriate setting for use of the procedure developed in this paper. All major analysts have proceeded on the basis that linear models and the assumption of OLS are appropriate. The sample is large and so too is the full set of independent variables. Moreover, attention has focused primarily on two major types of variables; those measuring characteristics associated with schools, and those measuring family and other social background characteristics. While the initial report set off a major series of critiques, rebuttals and re-analyses (see, esp., Mosteller and Moynihan, 1972 and Cain and Watts, 1970) which raised many kinds of theoretical, statistical, and policy related debates, most can be traced to one finding.

As Mostellar and Moynihan (1972, p. 8, quoted in Cnudde, 1974, p. 1) put it:

The pathbreaking quality of EEOR had to do with its analysis of the relation of variation in school facilities to variation in levels of academic achievement. It reported so little relation as to make it almost possible to say there was none

In particular, school differences exerted little impact independent of family and social background measures. Any apparent effects were substantially attributable to the collinearity between the two sets of variables. Note further that little attention was focused on any individual independent variable.

Cnudde's re-analysis examined four measures of school achievement; reading comprehension, verbal and non-verbal scores, and mathematical achievement. The dependent variables here will be mathematical achievement and reading

comprehension. We should note that these two variables seem the most likely to be measured with little error (including response-set) and also the most likely to be significantly influenced by the school characteristics (for details, see Cnudde, 1974, esp. p.8). Cnudde examined school summary variables (i.e., used the mean response for each school to each category of a given variable in the original measures). Thus, he focuses on differences across schools and school districts. Further specifics may be found in his analyses. Finally, note that we will limit our attention to the ninth and twelfth grades.

Table 3 reports some summary information about the estimation of the first step in the procedure (eq. 12, the interested reader is referred to Cnudde's analyses for the individual coefficient estimates). As noted earlier, these are the appropriate summary measures for the procedure and are unaffected by any multicollinearity. In all four cases, i.e., for the ninth and twelfth grades and for reading comprehension and mathematical aptitude, the variations across school districts are well explained (i.e., high $\bar{R}^2$'s and F's with significance easily surpassing the .0005 level). Indeed, the lowest corrected R^2 is .915 for mathematical aptitude among seniors. If we were to raise the question of the impact of specific variables or types of variables, we would immediately be overwhelmed by their large number. The number of independent variables ranges from 78 to 91. This number is exacerbated since we are here analyzing aggregate (i.e., school level) data. For example, to measure the affect of siblings requires a number of variables. At the individual level, number of siblings can be straightforwardly measured in a single variable. At the aggregate level, it requires a variable, "percentage of students with no siblings," another variable, "percentage of students with one sibling," etc. The increase in the number of independent variables is, at best, bothersome. However, the procedures necessitated by moving to the school level result in substantial increases in multicollinearity, both because aggregate data are often more highly related than individual data, and because the several

measures of, say, number of siblings will be correlated with each other. Finally, there will be even less interest in any one specific variable than on sets or types of independent variables. Therefore, we applied our technique to the first stage equations summarized in Table 1.

[Table 1 About Here]

Note that the first stage (which we assume to be correctly and fully specified) entails no assumptions beyond those of OLS. After all, the first step is OLS. Therefore, the results of this set of estimates can be used to define any partitioning of the variables we might choose. For example, $\hat{V}_1$ might be defined as the weighted summation of all sibling variables, and so forth. In this case, we take the partition that has been at the base of the "Coleman report" controversy; the division of variables into background characteristics (or $\hat{F}$ in the tables for "family and related background characteristic") and school factors (or $\hat{S}$). The results of the second stage, so defined, for the four regressions are found in Table 2.

[Table 2 About Here]

Attention focuses first on the question of significance. In this case both variables are "very" significant (except the .0005 level) in all equations. Note what we have concluded. In opposition to the summary statement of Mosteller and Moynihan cited earlier, variations in school facilities exerts significant impact on variations in academic achievement (albiet at the school level and for reading and mathematics) even controlling for the impact of variations in background characteristics. This statistical significance is, of course, reflected in the small standard errors of the instrumental variables.

While both variables exert a statistically significant impact on achievement, could it be that the substantive significance is rather small? After all, we have fairly large samples (N's of 922 and 778 for ninth and twelfth grades, respectively) and we have "pooled" the effects of a large number of

individual variables. Therefore, statistical significance, even very clear significance, may be misleading, and what is really interesting is the substantive impact of the two types of variables. Of course, there is no unambiguous and unproblematic measure of "substantive impact." The standardized beta is, however, often used as an indication of "importance." Recognizing its limitations, the results seem clear in this case. Both instrumental variables appear to exert a substantial impact on academic achievement. The pattern is interesting. The school characteristics seem to be relatively more consequential in predicting variations in reading comprehension as compared to mathematical aptitude (which might be expected if the mathematical variable is aptitude) and seem to be more important in twelfth than in ninth grade (which also might be expected if the effect of the school is cumulative over the 12 year period, for more details, see Cnudde, 1976). Thus, even with remaining multicollinearity, both types of variables exert a substantively as well as statistically significant impact on achievement. Interestingly, the standardized beta for the school instrumental variable in the twelfth grade reading equation is almost twice the size of that for background characteristics. It may well be that the effects of, say, school characteristics were masked due to the substantial repetitiveness and redundancy of individual measures.

Finally, we report briefly the results of the aforementioned procedure to measure the consequences of remaining collinearity (i.e., regress the dependent variable on each instrument individually). The results may be found in Table 3. Note that each variable is very strongly related to the dependent variable, and there is clearly a substantial amount of collinearity between background and school factors. Since each variable is so strongly related to Y, the extent to which the slope exceeds one is a direct measure of the extent to which the

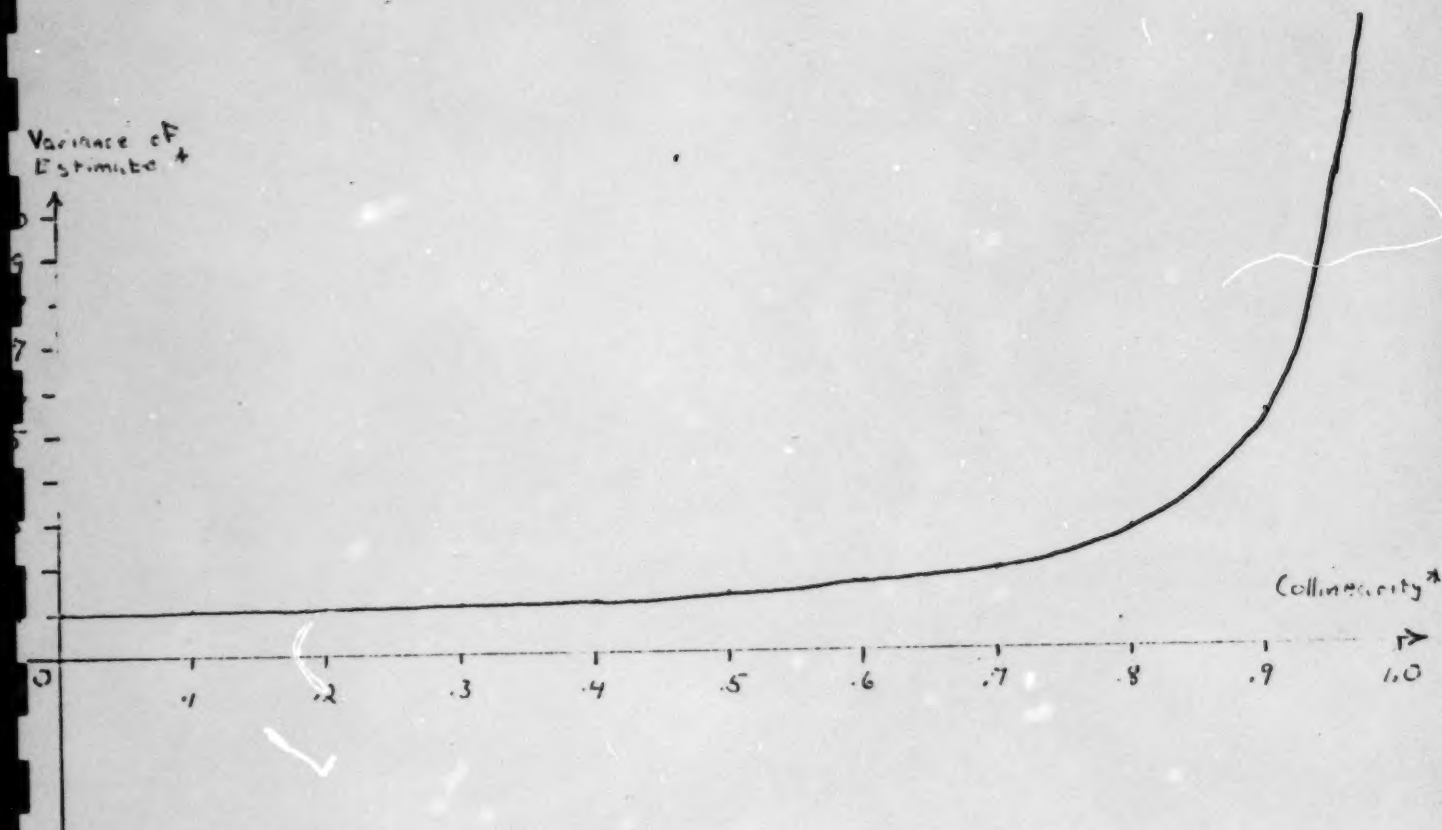
effect of the variable is dampened due to collinearity. Clearly, the hypothesis that the impact of any type of variable is not affected by remaining multicollinearity can be decisively rejected, and yet each type of variable exerts its own substantial and independent impact on academic achievement.

[Table 3 About Here]

This example application nicely illustrates some of the advantages of the proposed technique. In this situation, we were faced with a classic case of repetitiveness and redundancy. Moreover, there was a clear, a priori justification for partitioning the independent variables. And, in retrospect it seems that the conclusions others have reached, especially about the effects of school factors on achievement may well have been as they were not due to the lack of "true" effect of the school but due to the presence of multicollinearity.

Figure 1

Variance of Estimate as a Function
of Collinearity in Trivariate Case

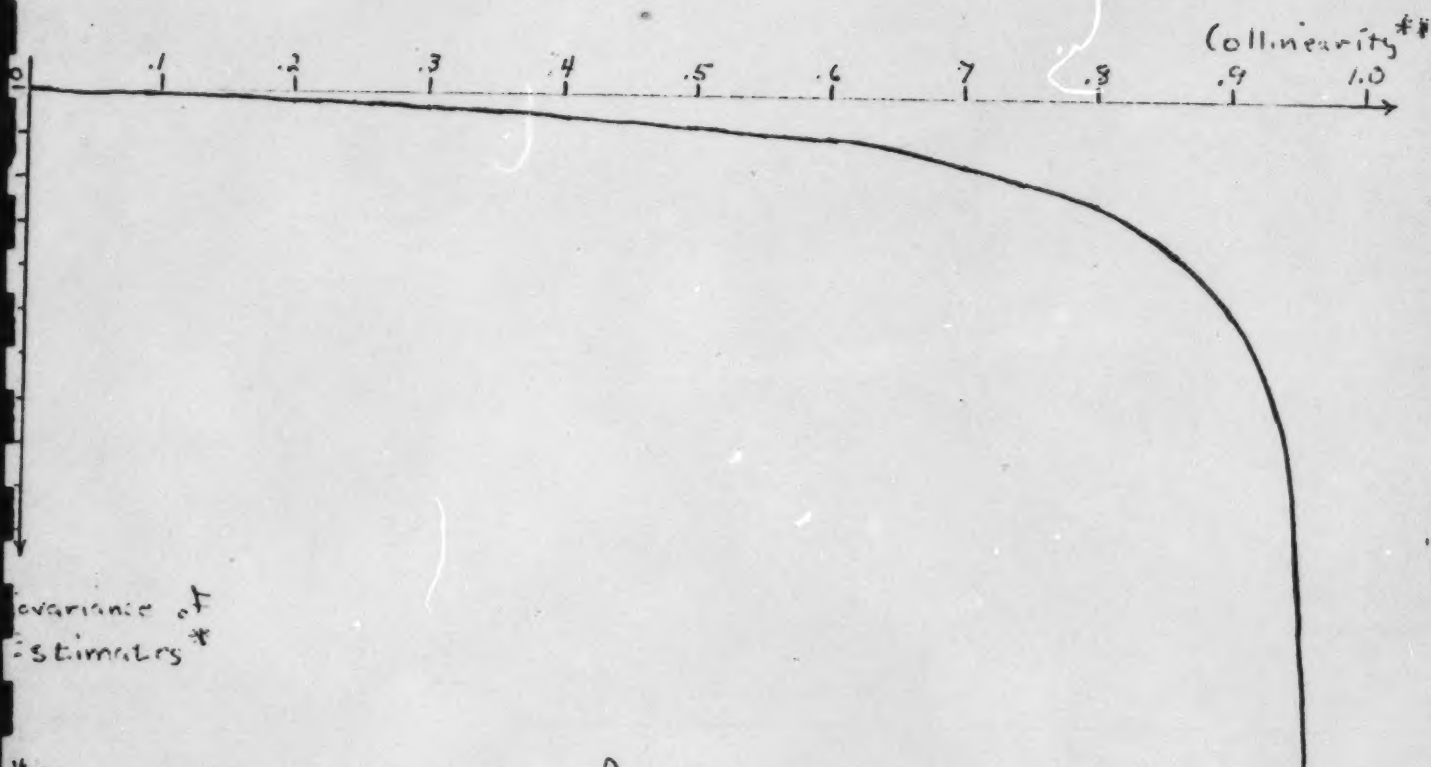


*From eq 5 in text with $\sigma^2 = 1$

**Pearson's Correlation Coefficient, from eq 5 in text

Figure 2

Covariance of Estimates as a
Function of Collinearity in the
Trivariate Case



Covariance of
Estimates*

*From solving eq. 6 in text with $\sigma^2 = 1$

**Pearson's Correlation Coefficient, from eq. 6 in text

Table 1
Summary Statistics of First Stage Regressions

Ninth Grade: Reading Comprehension: N = 922

Number of Background Variables = 50

Number of School Variables = 41

$$\overline{R}^2 = .933 \quad F(91,830) = 127 \quad \hat{\sigma}^2 = 3.07$$

Mathematical Aptitude: N = 922

Number of Background Variables = 47

Number of School Variables = 31

$$\overline{R}^2 = .924 \quad F(78,843) = 131 \quad \hat{\sigma}^2 = 2.52$$

Twelfth Grade: Reading Comprehension: N = 778

Number of Background Variables = 39

Number of School Variables = 41

$$\overline{R}^2 = .946 \quad F(80,697) = 154 \quad \hat{\sigma}^2 = 2.96$$

Mathematical Aptitude: N = 778

Number of Background Variables = 37

Number of School Variables = 53

$$\overline{R}^2 = .925 \quad F(90,687) = 93.6 \quad \hat{\sigma}^2 = 3.16$$

Table 2
Instrumental Variable Estimates of Academic Achievement

<u>Ninth Grade:</u>	<u>Reading Comprehension</u> =	$\hat{F}$	+	$\hat{S}$	
		(.031)		(.032)	Standard Error of Estimate
		.510		.498	Standardized Beta
		32.4		31.6	T
	<u>Mathematical Aptitude</u> =	$\hat{F}$	+	$\hat{S}$	
		(.029)		(.047)	Standard Error of Estimate
		.610		.382	Standardized Beta
		33.9		21.3	T
<u>Twelfth Grade:</u>	<u>Reading Comprehension</u> =	$\hat{F}$	+	$\hat{S}$	
		(.054)		(.031)	Standard Error of Estimate
		.352		.633	Standardized Beta
		18.5		32.2	T
	<u>Mathematical Aptitude</u> =	$\hat{F}$	+	$\hat{S}$	
		(.039)		(.045)	Standard Error Estimate
		.531		.460	Standardized Beta
		25.5		22.1	T

Table 3
Regressions of Each Instrument and Academic Achievement

<u>Ninth Grade:</u>	<u>Reading</u> = 1.82 $\hat{F}$ (.024) .928 75.2	<u>Reading</u> = 1.86 $\hat{S}$ (.025) .926 74.2	Standard Error of Estimate Standardized Beta T
	<u>Mathematics</u> = 1.54 $\hat{F}$ (.018) .942 84.7	<u>Mathematics</u> = 2.37 $\hat{S}$ (.036) .910 66.8	Standard Error of Estimate Standardized Beta T
<u>Twelfth Grade:</u>	<u>Reading</u> = 2.58 $\hat{F}$ (.035) .935 73.6	<u>Reading</u> = 1.52 $\hat{S}$ (.016) .961 96.3	Standard Error of Estimate Standardized Beta T
	<u>Mathematics</u> = 1.76 (.024) .937 74.4	<u>Mathematics</u> = 2.02 (.029) .928 69.4	Standard Error of Estimate Standardized Beta T